THE UNIVERSITY *of* EDINBURGH

Scaling Up Virtual MIMO Systems

Miryam Guadalupe Gonzalez Perez

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
September 2017

Abstract

Multiple-input multiple-output (MIMO) systems are a mature technology that has been incorporated into current wireless broadband standards to improve the channel capacity and link reliability. Nevertheless, due to the continuous increasing demand for wireless data traffic new strategies are to be adopted. Very large MIMO antenna arrays represents a paradigm shift in terms of theory and implementation, where the use of tens or hundreds of antennas provides significant improvements in throughput and radiated energy efficiency compared to single antennas setups. Since design constraints limit the number of usable antennas, virtual systems can be seen as a promising technique due to their ability to mimic and exploit the gains of multi-antenna systems by means of wireless cooperation. Considering these arguments, in this work, energy efficient coding and network design for large virtual MIMO systems are presented.

Firstly, a cooperative virtual MIMO (V-MIMO) system that uses a large multi-antenna transmitter and implements compress-and-forward (CF) relay cooperation is investigated. Since constructing a reliable codebook is the most computationally complex task performed by the relay nodes in CF cooperation, reduced complexity quantisation techniques are introduced. The analysis is focused on the block error probability (BLER) and the computational complexity for the uniform scalar quantiser (U-SQ) and the Lloyd-Max algorithm (LM-SQ). Numerical results show that the LM-SQ is simpler to design and can achieve a BLER performance comparable to the optimal vector quantiser. Furthermore, due to its low complexity, U-SQ could be consider particularly suitable for very large wireless systems.

Even though very large MIMO systems enhance the spectral efficiency of wireless networks, this comes at the expense of linearly increasing the power consumption due to the use of multiple radio frequency chains to support the antennas. Thus, the energy efficiency and throughput of the cooperative V-MIMO system are analysed and the impact of the imperfect channel state information (CSI) on the system's performance is studied. Finally, a power allocation algorithm is implemented to reduce the total power consumption. Simulation results show that wireless cooperation between users is more energy efficient than using a high modulation order transmission and that the larger the number of transmit antennas the lower the impact of the imperfect CSI on the system's performance.

Finally, the application of cooperative systems is extended to wireless self-backhauling heterogeneous networks, where the decode-and-forward (DF) protocol is employed to provide a cost-effective and reliable backhaul. The associated trade-offs for a heterogeneous network with inhomogeneous user distributions are investigated through the use of sleeping strategies. Three different policies for switching-off base stations are considered: random, load-based and greedy algorithms. The probability of coverage for the random and load-based sleeping policies is derived. Moreover, an energy efficient base station deployment and operation approach is presented. Numerical results show that the average number of base stations required to support the traffic load at peak-time can be reduced by using the greedy algorithm for base station deployment and that highly clustered networks exhibit a smaller average serving distance and thus, a better probability of coverage.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the School of Engineering at The University of Edinburgh.

Miryam Guadalupe Gonzalez Perez

Acknowledgements

First and foremost, I would like to express my deepest gratitude and appreciation to my supervisor Prof. John Thompson for his insightful technical advice and guidance throughout my studies. His close involvement and unfailing patience have been invaluable for my research.

I would like to thank my sponsor, the Mexican National Council for Science and Technology (Consejo Nacional de Ciencia y Tecnologia - CONACYT), which financial support gave me the opportunity to start and complete my doctoral studies.

I would like to offer my heartfelt thanks to my family. Your love, patience, understanding and support are worth far more to me than I could ever express.

A very special thank to Alex, whose daily encouragement and unconditional support have made this journey a wonderful experience.

I am also grateful to Dr. Martin Luna for his incredible support since the beginning of my career. Working with him has been a professional enriching experience.

Last but not the least, I would like to extend my sincere thanks to Rodrigo, Alessandro, Mohammad, Shendi, Kimin, Hamed and Jinho at the IDCoM for their continuous support and friendship.

Contents

List of figures

List of tables

Acronyms and abbreviations

4G	Fourth-generation
5G	Fifth-generation
A-BS	Anchored base station
AF	Amplify-and-forward
AoA	Angle of arrival
AoD	Angle of departure
ASE	Area spectral efficiency
AWGN	Additive white Gaussian noise
BICM	Bit-interleaved coded modulation
BLER	Block error rate
BS	Base station
CAGR	Compound annual growth rate
CCDF	Complementary cumulative distribution function
CDF	Cumulative density function
CF	Compress-and-forward
CoV	Coefficient of variation
CQ-PNC	Channel quantised physical-layer network coding
CSI	Channel state information
DF	Decode-and-forward
DL	Downlink
DPC	Dirty-paper coding
ECR	Energy consumption ratio
EE	Energy efficiency
GLA	Generalized Lloyd algorithm
GSM	Global system for mobile communications
HetNet	Heterogeneous network
HetHetNet	Heterogeneous HetNet
IAP	Improved affinity propagation
LBG	Linde-Buzo-Gray algorithm

LLR	Log-likelihood ratio
LM-SQ	Lloyd-Max scalar quantiser
LoS	Line of sight
LTE	Long-term evolution
LTE-A	Long-term evolution advanced
MISO	Multiple-input single-output
MIMO	Multiple-input multiple-output
ML	Maximum Likelihood
mm-wave	Millimetre wave
MSE	Mean square error
MMSE	Minimum mean-square error
MU-MIMO	Multi-user multiple-input multiple-output
NLoS	Non-line of sight
OFDM	Orthogonal frequency-division multiplexing
PA	Power amplifier
PCA	Principal component analysis
PDF	Probability density function
PP	Point process
PPP	Poisson point process
QAM	Quadrature modulation
QoS	Quality of service
RF	Radio frequency
SE	Spectral efficiency
SIC	Successive interference cancellation
SIMO	Single-input multiple-output
SINR	Signal-to-interference and noise ratio
SIR	Signal-to-interference ratio
SISO	Single-input single-output
SNR	Signal-to-noise ratio
STBC	Space-time block coding
SVQ	Sphere vector quantisation
TDD	Time division duplex
TDMA	Time-division multiple-access

TSVQ	Tree-structure vector quantiser
UDN	Ultra dense network
UE	User terminal
UIL	User-in-the-loop
UL	Uplink
ULA	Uniform linear array
UMTS	Universal mobile telecommunications service
U-SQ	Uniform scalar quantisation
USQV	Unitary space vector quantisation
V-MIMO	Virtual multiple-input multiple-output
VQ	Vector quantiser
V-VQ	Voronoi vector quantisation
WiFi	Wireless fidelity
WZ	Wyner-Ziv coding
ZF	Zero forcing

Nomenclature

δ	Efficiency of the power amplifier
η_a	Fraction of the bandwidth available for the users associated with the tagged A-BS
η_b	Fraction of the bandwidth available for all the associated BSs at an A-BS
$\eta_{ECR}(\mathcal{B})$	Sum ECR that can be achieved by the set of of BSs, $\mathcal{B}$
$\eta_{EE}(\mathcal{B})$	Sum EE that can be achieved by the set of of BSs, $\mathcal{B}$
ε	Channel estimation error
$\theta_{k,l}^{j}$	Vertical angle of arrival ($j = r$) or departure ($j = t$) from the l-th sub-path of the k-th cluster
κ_0	Average number of users correlated in location with the base station
κ_1	Average number of independent users served by the base station
λ_1	Density of the BS PPP
$\lambda^{(UE)}$	Density of the UE PPP
$\phi_{k,l}^{j}$	Horizontal angle of arrival ($j = r$) or departure ($j = t$) from the l-th sub-path of the k-th cluster
$\Phi_j^{(BS)}$	PPP formed by the locations of the BSs
$\Phi^{(UE)}$	PPP formed by the locations of the UEs
Φ_1^u	Cluster point process
χ_i^b	Set of symbols in the M-QAM constellation
σ_N^2	Noise power
σ_{CF}^2	Lower bound of the compress noise
τ_u	Number of symbols used for uplink pilots
$\psi_{k,l}$	Channel gain on the l-th sub-path of the k-th cluster
ω	Fraction of the BSs that have wired connection to the core of the network
$\mathbf{a}_r$	Antenna response vector at the receiver
$\mathbf{a}_t$	Antenna response vector at the transmitter
$\mathcal{B}$	Set of micro base stations
$\mathcal{B}_c$	Set of candidate micro base stations
$\mathcal{B}_M$	Set of macro base stations

$\mathcal{B}_{op}$	Set of micro base stations with wired connection to the core network
C_b	Codebook size
C	Capacity of the error-free conference link
C_{MIMO}	Ergodic capacity of MIMO systems
C_{MISO}	Ergodic capacity of MISO systems
$C_{MU-MIMO}$	Ergodic capacity of MU-MIMO systems
C_{SIMO}	Ergodic capacity of SIMO systems
C_{ZF}	Lower bound of the ergodic capacity of a MIMO system with ZF precoding
$\mathcal{D}_q$	Encoding region associated with the codevector v_q.
E_b	Energy per bit
E_s	Energy per symbol
$f_{a,k}$	Flag that indicates if a base station is used for access or not
$f_{b,k}$	Flag that indicates if a base station is used for backhaul or not
$f_{i,k}$	Flag that indicates if a base station is active or not
$f_{w,k}$	Flag that indicates the presence or absence of a wired connection from the base station to the core of the network
f_{bw}	Fraction of radio resources used for backhauling
$\mathbf{F}$	$N_t \times N_s$ precoder matrix
g	Relative channel strength of the relay and receiver node
G_B	Base station antenna gain
G_U	User antenna gain
h_{ij}	Channel coefficient from the j-th transmit antenna to the i-th receive antenna
h_{rd}	Channel coefficient from the relay to the destination
h_{sd}	Channel coefficient from the source to the destination
h_{sr}	Channel coefficient from the source to the relay
$\mathbf{H}$	$N_r \times N_t$ channel matrix
$\hat{\mathbf{H}}$	MMSE estimate of the channel matrix
$\mathcal{I}_0$	Interference generated by the cluster centre
$\mathcal{I}_j$	Interference generated by all the other BSs excluding the tagged BS when the typical user is served by its cluster centre ($j = H$) or not ($j = 1$)
I_a	Interference in the access link
I_b	Interference in the backhaul link
I_s	Number of iterations to meet the stopping criterion of codebook design

L	Number of symbol periods
m_s^b	Number of bits transmitted per data symbol
M	Number of points on the constellation
$\mathbf{n}$	Noise vector
N_0	Noise power
N_b	Number of base stations with wireless backhaul links to the tagged A-BS
N_c	Number of relay nodes
N_d	Number of destination nodes
N_l	Number of scatterers
N_p	Number of points per dimension
N_{rb}	Number of resource blocks
N_r	Number of antennas at the receiver
N_s	Number of data streams
N_t	Number of antennas at the transmitter
N_u	Number of users served by the tagged BS
$N_{u,w}$	Number of users served by the tagged A-BS
$O_{h,\mathrm{LM}}$	Overhead bits incurred by the LM-SQ algorithm
$O_{h,\mathrm{USQ}}$	Overhead bits incurred by the U-SQ algorithm
$O_{h,\mathrm{VQ}}$	Overhead bits incurred by the V-SQ algorithm
p	Percentage of inhomogeneity of the network
P_0	Load-independent power consumption
P_{BS}	Power consumption at the BS
$\mathbb{P}_{C_j}$	Probability of coverage of the homogeneous ($j = H$) or inhomogeneous ($j = T$) user distribution
P_{cod}	Power consumption for coding
P_{cons}	Total power consumption of the system
P_{cj}	Circuit power consumption at the transmitter ($j = t$) or receiver ($j = r$)
P_{dec}	Power consumption for decoding
$\mathbb{P}_{e,\mathrm{c}}$	Block error probability at a destination node
$\mathbb{P}_{e,\mathrm{d}}$	Block error probability at a relay node
P	Transmit power of the BS
P_a	Transmit power for access
P_b	Transmit power for backhaul

P_i	Transmit power of the source ($i = s$) or relay ($i = r$)
P_{lo}	Power consumed by the local oscillator
P_M	Transmit power of a macro BS
P_j	Power consumption of the network when the random ($j = RS$) and load-based ($j = LB$) sleeping algorithms are used
P_{pr}	Power consumption for signal processing
P_{se}	Power consumption of the system once some BSs have been switched-off
P_{UE}	Power consumption at the receiver side
q	Probability that a base station remains active
R_b	Code rate
R_{bh}	Achievable rate of the backhaul links
R_c	Source coding rate
R_{fh}	Achievable rate of the access links
R	Achievable rate
$\mathbf{R}$	Noise covariance matrix
R_{AF}	Achievable rate for the AF protocol
R_{DF}	Achievable rate for the DF protocol
R_{CF}	Achievable rate for the CF protocol
s_k	Symbol transmitted to the k-th user
$\hat{s}_k$	Estimate of s_k
$\mathbf{s}$	$N_s \times 1$ vector of data symbols
$\mathcal{S}$	Set of data symbols of size $\lvert\mathcal{S}\rvert = 2^{m_s^b}$
T	Throughput of the network
$T_{d,\text{LM}}$	Design time for the Lloyd-Max algorithm
$T_{d,\text{USQ}}$	Design time for the uniform scalar quantisation
$T_{d,\text{VQ}}$	Design time for the vector quantisation algorithm
T_c	Computational time for comparing the distortion between two vectors
T_d	Computational time for finding the distance between two vectors
T_{LBG}	Computational time for the Linde-Buzo-Gray algorithm
v	Carrier Frequency
v_q	Codevector associated with the encoding region $\mathcal{D}_q$.
V_s	Number of training vectors used for codebook design
W	Bandwidth of the system

$\mathbf{W}$	$N_r \times N_s$ combining matrix
$x_{s,i}$	Transmit signal from the source during the i-th time slot
$\mathbf{x}$	Transmit signal vector
$\hat{\mathbf{x}}_{\text{ZF}}$	Equalized signal at the ZF demodulator
$\mathbf{y}$	Received signal vector
y_c^{NF}	Noise-free constellation of the received signal at the relay
y_k	Received signal at the k-th user
y_r	Received signal at the relay
$y_{d,i}$	Received signal at the destination during the i-th time slot
${}_2\mathcal{F}_1[\cdot]$	Gaussian hypergeometric function
$\mathcal{CN}(\mu, \sigma^2)$	Complex Gaussian distribution with mean μ and variance σ^2
$\exp(\mu)$	Exponential distribution with mean μ
$\mathbb{E}[\cdot]$	Expected value operator
$Im\{\cdot\}$	Imaginary part operation
$\mathcal{L}_x(\cdot)$	Laplace transform of x
$\mathbb{P}_X(x)$	Probability that the event x occurs
$Re\{\cdot\}$	Real part operation
$\lvert\mathcal{X}\rvert$	Cardinality of the set $\mathcal{X}$
$(\cdot)^*$	Matrix conjugate operation
$(\cdot)^H$	Matrix conjugate-transpose operation
$(\cdot)^T$	Matrix transpose operation
$\bar{x}$	NOT Boolean operator
$\sup(\cdot)$	Supremum operator

Chapter 1
Introduction

This thesis focuses on the design of energy-efficient multiple-input multiple-out systems with very large antenna arrays and its integration with the concept of cooperative wireless systems. The origin and motivation of this work are provided in Section 1.1. Followed by the overview of the organisation of the remaining chapters in Section 1.2. Finally, the main contributions of this thesis are summarised in Section 1.3.

1.1 Motivation

In recent years, demand for cellular data has been growing considerably and it is estimated that the traffic generated by smart-phones will increase by 10 times between 2017 and 2023 with a compound annual growth rate (CAGR) of 42% [6]. Current fourth-generation (4G) systems use advanced technologies such as orthogonal frequency-division multiplexing (OFDM), multiple-input multiple-output (MIMO) and turbo codes in order to achieve spectral efficiencies close to theoretical limits [2].

Since the design of current wireless networks have been mainly driven by enabling high spectral efficiency, these networks are characterised by poor energy efficiency (EE). With a total number of mobile subscribers of more than half of the global population, the telecommunications industry is becoming a major consumer of global energy. Thus, in recent years a great interest from academia and industry has arisen in technologies that aim at designing wireless networks with better coverage and higher EE [7]. The motivation is twofold. Firstly, the telecommunication network operators are experiencing significant impact on their cash flows due to energy cost. Secondly, the social responsibility of environmental protection by reducing the carbon footprint due to information and communications technology has become very relevant [4]. The ultimate goal is, thus, to design network architectures and technologies needed to meet the explosive growth in cellular data demand without increasing the overall energy consumption.

Base station sleeping strategies, relay and cooperative communications, resource allocation techniques and novel radio access techniques have been used as potential solutions for energy

efficient wireless communication systems [4, 8]. By using picocells and femtocells, network deployment strategies have been considered as an alternative to improve the energy efficiency of the network by decreasing the average path loss [8]. Recently, research on base station energy saving has also gathered great interest since the energy consumption breakdown over the mobile network elements illustrates that, although the core and data centres are the most energy greedy, up to 80% of the overall energy is consumed by the base stations [9]. To reduce energy consumption, China Mobile have deployed a dynamic power algorithm saving up to 27% of the total energy consumption [10]. Moreover, a scalable base station sleeping strategy complemented by the use of cooperative communications and power control algorithm was proposed in [11].

The energy efficiency, defined as the number of bits transferred per Joule of energy, is affected by many factors such as spectral efficiency, radiated transmit power, and circuit power consumption [12]. Generally speaking, the design of energy efficient networks can be achieved by using specifically designed architectures, radio technologies and resource management techniques. Some of the techniques that have been identified as potential solutions for the fifth-generation (5G) mobile systems are listed in Table 1.1.

Energy-Efficient:	Solutions and Features
Architectures	- Optimization of the cell size: large vs small cell deployment - Overlay resource: microcell, picocell or femtocell - Relay and cooperative communications
Resource Management	- Power and resource allocation - Single-input single-output (SISO) vs MIMO scheduling
Radio Technologies	- Heterogeneous Network deployment - Millimetre wave communications - Simultaneous wireless information and power transfer

Table 1.1: Techniques for energy efficiency [4].

Though relay and cooperative communications were initially introduced as an alternative to increase coverage area, later these technologies were also used to increase the throughput as well as decrease the transmission power [1,13,14]. In the Long-Term Evolution advanced (LTE-A) standard, in-band and out-band relaying standards have been included. Nevertheless, some

issues like full duplex relays and connectivity with mobile relays are left for implementation in the next generation mobile networks, where it is expected that different relay policies will be supported.

1.2 Thesis Outline

The use of emergent techniques such as Massive MIMO, millimetre-wave frequencies as well as the use of small cells and relays in the next generation networks impose new research challenges. This work is aimed at designing a multi-user MIMO system for optimal energy efficiency. In particular, bringing insights on how relay and cooperative techniques can be integrated with very large antenna MIMO systems. Moreover, since 5G systems are expected to support heterogeneous networks, the trade-offs between spectral efficiency, probability of coverage and energy efficiency are investigated.

1.2.1 Background

Chapter 2 provides a comprehensive basis for the research presented in Chapters 3-5. First, some fundamental principles of MIMO systems are provided. A study of the system's performance when the number of antennas employed is increased will be illustrated. Then, cooperative wireless systems are presented as a practical alternative to traditional MIMO systems. Thus, a general overview of the benefits of using Virtual MIMO systems are described. In addition, an introduction to millimetre-wave communication systems is given. Finally, the principles of stochastic geometry for heterogeneous networks are presented.

1.2.2 Low Complexity Codebook Design

Chapter 3 considers a cooperative virtual MIMO (V-MIMO) system that uses a large multi-antenna transmitter and implements bit-interleaved coded modulation (BICM) transmission and compress-and-forward (CF) relay cooperation among multiple receiving nodes. Since constructing a reliable codebook is the most computationally complex task performed by the relay nodes in CF cooperation, reduced complexity quantisation techniques are introduced in this chapter. In order to focus the analysis on the block error probability (BLER) and computational complexity, an error-free conference link between them is considered. Two different scalar

quantisation techniques are presented, where the real and imaginary parts of the received signals are quantised independently. First, the uniform scalar quantiser (U-SQ) is considered and the performance achieved when a large number of antennas is used both at the transmitter and receiver is evaluated. Second, the Lloyd-Max algorithm (LM-SQ) is employed to reduce the error distortion by taking advantage of the distribution of the received signals. A comparison in terms of BLER and computational complexity for designing the codebook and forwarding a compressed version of the signal to the destination node are presented. In practice, besides the data symbols, additional control information is transmitted, thus, the overhead incurred by the quantisation techniques is computed by calculating the number of bits required to transmit the codebook to the destination node.

1.2.3 Energy Efficient Massive MIMO Wireless Networks

In Chapter 4, a cooperative Virtual multi-user MIMO (MU-MIMO) system that uses a large multi-antenna transmitter and implements BICM transmission and decode-and-forward (DF) relay cooperation among multiple receiving nodes is considered. In DF-based cooperation, decoding, re-encoding and forwarding the estimated received signals to the destination node for subsequent processing are the main tasks that each relay has to perform. Even though large multi-antenna approaches have been pointed as an effective way to enhance the spectral efficiency (SE) of wireless networks, this comes at the expense of linearly increasing the power consumption due to the use of multiple radio frequency (RF) chains to support the antennas. Thus, in this chapter, the energy efficiency and the effective throughput of the wireless system are analysed. In order to quantify any energy efficiency gains achieved by the cooperative wireless system a common framework is required. Thus, a power consumption model that considers not only the power consumed by the transmitter but also that takes into account the overhead power consumption incurred by the relays is introduced. Moreover, the impact of the imperfect channel state information (CSI) on the throughput and energy efficiency of the system is studied and a power allocation algorithm to reduce the total power consumption is implemented.

1.2.4 Energy Efficient Inhomogeneous Cellular Networks

In Chapter 5, a wireless self-backhauling heterogeneous network is studied. Thus, it is assumed that the macro-cell base station (BS) controls the user scheduling, resource allocation and sup-

ports high-mobility users, while micro BSs, with a smaller coverage area, provide high data rates for low-mobility users. Each micro base station is assumed to be equipped with large multi-antenna transmitter to perform backhaul and access operations using millimetre wave (mm-wave) frequencies. Furthermore, a DF cooperation algorithm is employed to provide a cost-effective and reliable backhaul. An energy efficient base station deployment and dynamic operation by using the heuristic greedy algorithm are presented. Due to the inhomogeneity of the traffic load in 5G heterogeneous networks (HetNets), first, it is assumed that the users are inhomogeneous distributed in space with the average number of users varying over time. Later, it is assumed that the average number of users to be served is fixed while their distribution in space is modelled by a Thomas Cluster Process [15].

1.2.5 Conclusions

Finally, conclusions of the previous chapters and future work are presented in Chapter 6.

1.3 Contributions

The main contributions for this research in cooperative wireless systems with large antenna arrays are described as follows:

- The Lloyd - Max (LM-SQ) algorithm and Uniform Scalar Quantisation (U-SQ) at each relay node are proposed as enablers for cooperative MIMO systems with large antenna arrays. An analysis of the design complexity and BLER of the wireless cooperative system is presented. It is shown that the LM-SQ is less complex and achieves a comparable performance to the optimal vector quantiser. Meanwhile, U-SQ is much simpler and exhibits a good performance when a large cooperative array is considered, making it suitable for these systems [16].

- A power consumption model that considers the overhead incurred by the relays is introduced. Using this model as a framework, an analysis of the energy efficiency and throughput for a virtual MIMO system with a large transmit antenna array is presented. Moreover, the impact of the imperfect CSI on the system's performance is studied and a power allocation algorithm is implemented to reduce the total power consumption. It is shown that when the number of available resource blocks is restricted, cooperative

systems can be employed to meet the demand for high data rates [17].

- Using an heuristic greedy algorithm, an energy-efficient base station deployment for mm-wave self-backhauling networks is presented. Moreover, a dynamic operation algorithm that allows the base stations to reduce power consumption by switching operation modes during low traffic load conditions is introduced. It is shown that the average number of additional micro base stations required by the greedy algorithm to support the traffic load at peak-time was 17.6% lower compared with a random selection [18].

- A tunable model that can cover a wide range of scenarios from a homogeneous Poisson point process to highly heterogeneous clustered wireless networks is introduced. Using the distribution of the serving distance, the probability of coverage and throughput of a wireless self-backhaul system with inhomogeneous user distribution are derived. Additionally an energy saving assessment using sleeping algorithms is performed by using random, load-based and greedy policies. It is shown that the inhomogeneous nature of the system provides an advantage for the algorithms where the load of the BSs and the resource allocation are considered [19].

Chapter 2
Background

In this chapter, an essential framework for the understanding of this thesis is provided. First, some fundamental principles of multiple-input multiple-output (MIMO) systems are given. A study of the system's performance when the number of antennas employed is increased will be illustrated. Then, cooperative wireless systems are presented as a practical alternative to traditional MIMO systems. Thus, a general overview of the benefits of using virtual MIMO systems are provided. In addition, an introduction to millimetre wave communication systems is given. Finally, the principles of stochastic geometry for heterogeneous networks are presented. The content of this chapter is therefore divided into four main sections: MIMO Wireless Systems, Cooperative Wireless Communications, Millimetre Wave Wireless Systems and Stochastic Geometry for Cellular Networks. Thereby, this chapter provides the reader with the essential knowledge of the state-of-the-art which will be frequently referred to in remainder of the thesis.

2.1 MIMO Wireless Systems

The use of multiple antennas at the transmitter and receiver in wireless systems, referred to as multiple-input multiple-output systems, gained attention over the past decade due their ability to improve both the channel capacity and the link reliability of wireless communications. In this section, the concept of MIMO systems, spatial multiplexing and spatial diversity are introduced. Furthermore, the extension of MIMO systems to the so called Massive MIMO concept is presented.

2.1.1 MIMO System Model

Consider a narrowband point-to-point MIMO communication system equipped with N_t antennas at the transmitter and N_r antennas at the receiver, as shown in Figure 2.1(a). Assuming frequency-flat block Rayleigh fading, where the channel matrix entries are constant during

each block but independently Rayleigh distributed on different blocks, the so-called quasi-static MIMO channel at a given time may be represented as

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_t} \\ h_{21} & h_{22} & \cdots & h_{2N_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r1} & h_{N_r2} & \cdots & h_{N_rN_t} \end{bmatrix}, \tag{2.1}$$

where the element h_{ij} is the fading coefficient from the j-th transmit antenna to the i-th receive antenna modelled as zero-mean circularly symmetric complex Gaussian random variable [20] with $\mathbb{E}[|h_{ij}|^2] = 1$. The input-output relation of the system is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.2}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_{N_r}]^T$ represents the received signal vector, $\mathbf{x} = [x_1, x_2, \ldots, x_{N_t}]^T$ is the transmitted signal vector and $\mathbf{n} = [n_1, n_2, \ldots, n_{N_r}]^T$ is the noise vector with entries $n_i \sim \mathcal{CN}(0, N_0)$.

Since the channel matrix at the receiver can be obtained by using training pilot signals [20], it is assumed that the channel state information (CSI) is known at the receiver and the transmitter side. Typically, this training information is sent to the transmitter via a feedback channel. With the knowledge of the CSI at the transmitter, optimal power allocation algorithms may be utilized to assign various levels of transmitted power to each of the transmit antennas. Due to the limitation on the feedback channel in practical wireless systems, it is not always possible to obtain CSI at the transmitter. If that is the case, the signals are assumed to be transmitted through each antenna with equal power level of E_s/N_t, where E_s is the total transmit power.

2.1.2 Benefits of MIMO Systems

The performance improvements achieved by MIMO systems are due to array gain, diversity gain and spatial multiplexing gain. In the following, each of these gains is briefly reviewed considering a system with N_t-transmit and N_r-receive antennas, frequently referred to as $N_t \times N_r$ MIMO system.

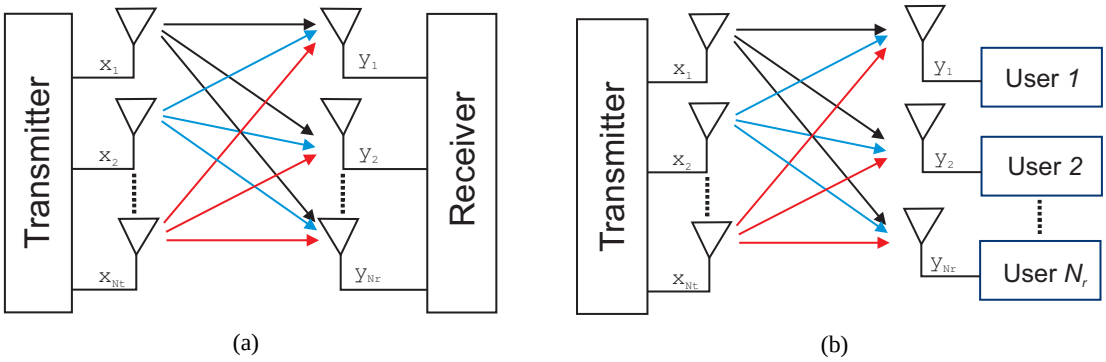


Figure 2.1: Schematic representation of a (a) point-to-point and (b) multi-user MIMO wireless system, with a N_t-antenna transmitter and N_r-antenna receiver or N_r single-antenna users, respectively.

2.1.2.1 Array Gain

Array gain is the increase in receive signal-to-noise-ratio (SNR) that results from the coherent combining of wireless signals at a receiver [21]. Coherent combining may be realized through spatial processing at the receive antenna array and/or spatial pre-processing at the transmit antenna array and thus requires channel knowledge in the transmitter and receiver, respectively, and depends on the number of transmit and receive antennas. Array gain improves the coverage of the wireless network by increasing the resistance of the system to noise [20, 21].

2.1.2.2 Diversity Gain

In a wireless system, the signal power at the receiver fluctuates randomly (or fades). Diversity is a technique that mitigates fading in wireless links by providing the receiver with multiple (ideally independent) copies of the transmitted signal in space, frequency or time [20, 21]. With an increasing number of independent copies, the probability that at least one of these copies does not experience a deep fade increases, thereby improving the quality and reliability of reception. A MIMO system equipped with N_t-transmit antennas and N_r-receive antennas could offer a maximum of $N_t \times N_r$ independently fading links, and hence a spatial diversity of order $N_t \times N_r$ [22].

2.1.2.3 Spatial Multiplexing Gain

MIMO systems offer a linear (in the number of transmit-receive antenna pairs) or $\min(N_t, N_r)$ increase in capacity for no additional power or bandwidth expenditure [23]. This gain, referred

to as spatial multiplexing gain is achieved by transmitting multiple, independent data signals from the individual antennas. Under suitable channel conditions, such as rich scattering, the receiver can separate these data streams. Considering that each receive antenna observes a superposition of the transmitted signal, various detector architectures, such as maximum likelihood (ML), zero-forcing (ZF), minimum mean-square error (MMSE), and successive interference cancellation (SIC) detector, can be used to achieve spatial multiplexing gain [24, 25]. In general, the number of data streams that can be reliably supported by a MIMO channel equals $\min(N_t, N_r)$ [26].

2.1.2.4 Capacity of Fading MIMO Channels

MIMO systems can offer substantial improvements over conventional single-input single-input (SISO) systems by employing diversity and spatial multiplexing techniques. In this section, the gain offered by MIMO in terms of system capacity is explored.

Assume an ergodic block fading channel model where the channel remains constant over a block of consecutive symbols and changes independently across blocks. The average SNR at each receive antenna is given by E_s/N_0 since $\mathbb{E}[|h_{ij}|^2] = 1$ in (2.2). If the transmitted symbols span an infinite number of independently fading blocks, the Shannon capacity, also known as ergodic capacity, is the ensemble average of the information rate over the distribution of the elements of the channel matrix $\mathbf{H}$ [27]. Assuming that the channel is known at the receiver and no channel state information at is available at the transmitter, the ergodic capacity can be expressed as [23, 28].

$$C_{MIMO} = \mathbb{E}\left[\log_2 \det\left(\mathbf{I}_{N_r} + \frac{E_s}{N_t N_0}\mathbf{H}\mathbf{H}^H\right)\right]. \tag{2.3}$$

where $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator. Single-input multiple-output (SIMO) and Multiple-input single-output (MISO) channels are specials cases of MIMO channels. Consider a SIMO channel $\mathbf{h}$ with $N_t = 1$ and N_r-receive antennas. Then, the capacity of the SIMO channel is given by

$$C_{SIMO} = \mathbb{E}\left[\log_2\left(1 + \frac{E_s}{N_0}||\mathbf{h}||^2\right)\right]. \tag{2.4}$$

On the other hand, if a MISO channel with N_t-transmit antennas and $N_r = 1$ is considered,

then the capacity of the MISO channel is given by

$$C_{MISO} = \mathbb{E}\left[\log_2\left(1 + \frac{E_s}{N_t N_0}||\mathbf{h}||^2\right)\right]. \tag{2.5}$$

Due to the inability of the MISO transmitter to exploit array gain when the channel is unknown at the transmitter, it is clear from equations (2.4) and (2.5) that $C_{SIMO} > C_{MISO}$. Using equation (2.1) to model the MIMO channel, Figure 2.2 shows the ergodic capacity of several MIMO configurations as a function of SNR. As expected, at the same SNR, the ergodic capacity increases with N_t and N_r. Since, MISO channels do not offer array gain in the absence of CSI at the transmitter, then the ergodic capacity of a 1×2 SIMO channel is greater than that of a 2×1 MISO channel as shown in Figure 2.2.

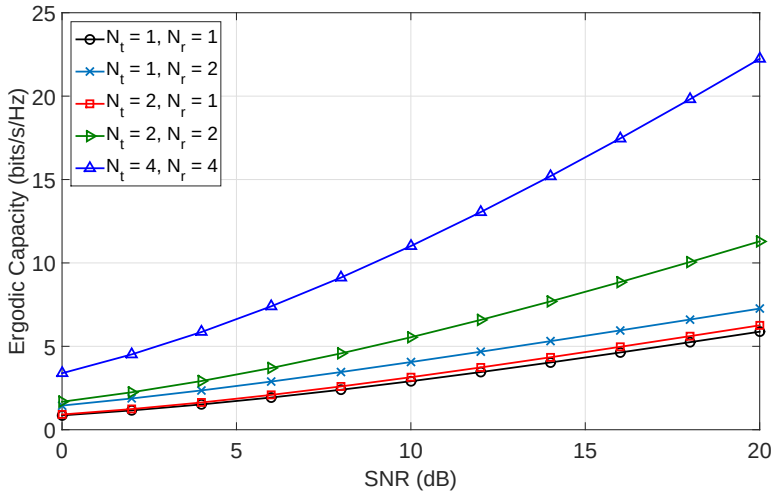


Figure 2.2: Ergodic capacity for different MIMO antenna configurations, when Rayleigh-fading channels are considered.

2.1.3 Multi-user MIMO

Aiming to exploit the spatial multiplexing gain, MIMO technology has shifted from a point-to-point to a multi-user approach, where several users are simultaneously served by a multiple-antenna transmitter (see Figure 2.1(b)). Multi-user MIMO (MU-MIMO) has two advantages over point-to-point MIMO. First, it is less vulnerable to the propagation environment, even under line-of-sight conditions. Second, only single-antenna terminals are required [29, 30].

Consider a MU-MIMO system which consists of a N_t-antenna transmitter and N_r single-antenna users. The transmitter sends the modulated data symbols using a transmit vector $\mathbf{x}$

that satisfies the power constraint $\mathbb{E}[||\mathbf{x}||^2] = E_s$ and linearly precodes the symbols using the channel estimate. Let s_k be the symbol to be transmitted to the k-th user, with $\mathbb{E}[|s_k|^2] = 1$. Then, the linearly precoded vector $\mathbf{x}$ can be written as

$$\mathbf{x} = \sqrt{p_d}\mathbf{F}\mathbf{s}, \tag{2.6}$$

where $\mathbf{s} \in \mathbb{C}^{N_r \times 1}$ is the vector of data symbols, $\mathbf{F} \in \mathbb{C}^{N_t \times N_r}$ is the precoding matrix and $p_d = E_s/\mathbb{E}[\text{tr}(\mathbf{F}\mathbf{F}^H)]$ is a normalization constant that satisfies $\mathbb{E}[||\mathbf{x}||^2] = E_s$. As a result, the received signal at the k-th user is given by

$$y_k = \sqrt{p_d}\mathbf{h}_k\mathbf{f}_k s_k + \sum_{j \neq k} \sqrt{p_d}\mathbf{h}_k\mathbf{f}_j s_j + n_k, \tag{2.7}$$

where $\mathbf{h}_k$ and $\mathbf{f}_k$ are the k-th row and column of the channel and precoding matrix, respectively. The Shannon sum-capacity for downlink MU-MIMO requires the solution of a convex optimization problem [29]

$$C_{MU-MIMO} = \sup_{\mathbf{a}} \{\log_2 \det\left(\mathbf{I}_{N_t} + p_d\mathbf{H}\mathbf{D}_a\mathbf{H}^H\right)\}, \tag{2.8}$$

with $\mathbf{a} \geq \mathbf{0}$, $\mathbf{1}^T\mathbf{a} = \mathbf{1}$, where $\mathbf{D}_a$ is a diagonal matrix whose diagonal elements comprise the vector $\mathbf{a} \in N_t \times 1$, and $\mathbf{1}$ denotes the $N_t \times 1$ vector of ones.

2.1.4 Very large antenna MIMO arrays

The performance of practical MIMO systems strongly depends on properties of the antenna arrays and the propagation environment, limiting the capacity of the system. What prevents the Shannon-theoretic version of MU-MIMO from being scalable is, first, the exponentially increasing complexity of coding and decoding, and second, the time spent acquiring CSI which increases with both the number of service antennas and the number of users [29]. When the number of antenna elements in the arrays increases, both opportunities and challenges arise. Massive MIMO is a MU-MIMO cellular system where the number of transmit antennas and the number users are large. In this section, the concept behind Massive MIMO systems in terms of communication schemes and signal detection is introduced.

2.1.4.1 Massive MIMO

In traditional MU-MIMO systems, precoding in the downlink and detection in the uplink require CSI at the transmitter. Thus, the time-frequency resources required for channel estimation are proportional to the number of transmit antennas and independent of the number of receive antennas. Generally speaking, two different operating modes can be used for transmission, frequency-division duplexing (FDD) and time-division duplexing (TDD). The downlink CSI acquisition in a FDD operation requires a two-stage procedure. The transmitter first sends out pilot waveforms, based on which the terminals then estimate the channel responses, quantise the estimates, and feed them back to the base station [31]. For systems with a large number of transmit antennas, such a procedure would reduce the spectral efficiency significantly. Instead, a common approach in Massive MIMO is to operate in TDD mode, and rely on the reciprocity of the channel to compute the pre-coding coefficients based on uplink CSI [32]. In practice, the uplink and downlink channels are not perfectly reciprocal due to mismatches of the RF chains [33]. Nevertheless, these mismatches can be removed by calibration [32, 34, 35]. Moreover, as the number of transmit antennas is typically increased to several times the number of users, simple linear precoders and decoders can be employed on the downlink (DL) and uplink (UL) to approach the Shannon limit [29].

Massive MIMO technology relies on the coherent but computationally very simple processing of signals from the antennas at the transmitter. Some benefits of a Massive MU-MIMO system are [29, 30, 33]: a) improvements of the capacity and simultaneously of the radiated energy efficiency, b) support for inexpensive low-power components, c) the reduction of latency on the air interface, d) the simplification of the multiple access layer and e) the increment on the robustness against interference and jamming. Nevertheless, Massive MIMO also exhibits some limiting factors as [29, 30]: a) pilot contamination, b) its performance relies on a favourable propagation channel, i.e., the channel responses from the transmitter to different user terminals are sufficiently different, c) the increment on the internal power consumption, d) increased hardware and baseband processing at the transmitter.

Figure 2.3 compares the Shannon spectral efficiency (black curve) and that achieved by Massive MIMO with linear precoding (blue curve) as a function of the number of antennas at the transmitter when $N_r = 16$ users and an average SNR $= 0$ dB are considered. The results in this plot are computed according to equation (2.8) for the Shannon limit, while the linear precoding

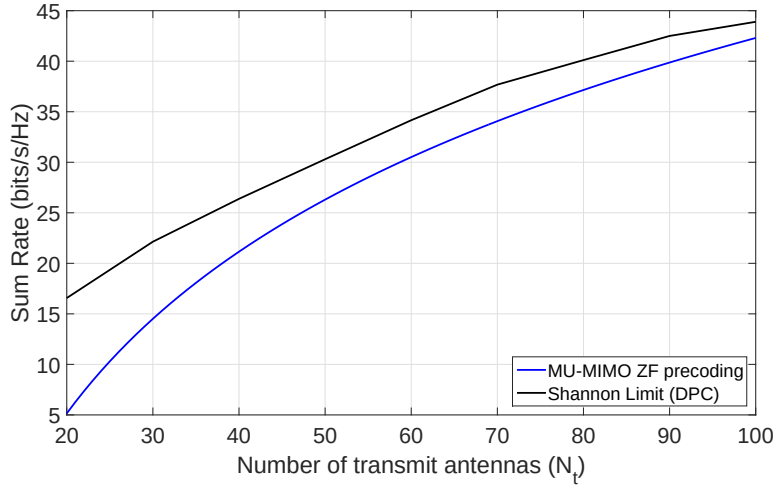


Figure 2.3: Comparison of the total spectral efficiency versus the number of antennas at the transmitter of the Shannon limit (dirty-paper coding) and Massive MIMO (linear pre-coding) systems for $N_r = 16$ users and SNR $= 0$ dB.

lower bound is computed according to the zero-forcing lower bound given in [36] as

$$C_{ZF} = N_r \log_2 \left(1 + \frac{(N_t - N_r)p_d}{N_r}\right). \tag{2.9}$$

If the transmitter is equipped with $N_t = 64$ antennas, then Massive MIMO provides the same performance as the optimal dirty-paper coding for $N_t = 55$, thus, by employing additional antennas at the transmitter, zero-forcing can offer competitive performance compared with the dirty-paper coding technique that is required to achieve Shannon-capacity in Massive MIMO.

2.2 Cooperative Wireless Communications

As it was mentioned in Section 2.1, MIMO systems have drawn significant interest due to their ability to increase both capacity and reliability of modern wireless communications. Nevertheless, design constraints for mobile devices limit the number of usable antennas; further, closely spaced antennas entail the presence of highly correlated fading signals, limiting performance [37]. Cooperative systems can be considered as a practical alternative to traditional MIMO systems when the size and cost of the wireless device is limited, allowing single-antenna devices in a multi-user scenario to share their messages [38]. In this section, the concept of cooperative wireless communications is introduced. First, some cooperation configurations and operation types are outlined, followed by an introduction of typical relaying schemes and a

comparison of their achievable rates. Finally, an introduction to virtual MIMO systems is provided.

2.2.1 Configurations and Operation Types

Consider the three-node classical relay channel shown in Figure 2.4(a), where the source node (S) transmits to the destination node (D) with the assistance of the relay node (R) [37], where the source terminal, the destination and relay are equipped with single-antenna transmitter and receivers. Moreover, it is assumed that the nodes operate in a half-duplex mode, and thus, cannot transmit and receive simultaneously. From the source's point of view, the system becomes a multiple-access channel. At the destination, the cooperative system can be seen as broadcast channel, with all receivers coordinating to perform joint detection [39].

The classical relay channel concept can be extended to larger network architectures. Figure 2.4 illustrates various relaying configurations that can be found in wireless networks, such as the parallel relay channel in Figure 2.4(b), the virtual MIMO (V-MIMO) channel in Figure 2.4(c), and the two-way relay channel in Figure 2.4(d) [24]. According to the scope of this work, the V-MIMO system will be presented in more detail in Section 2.2.3.

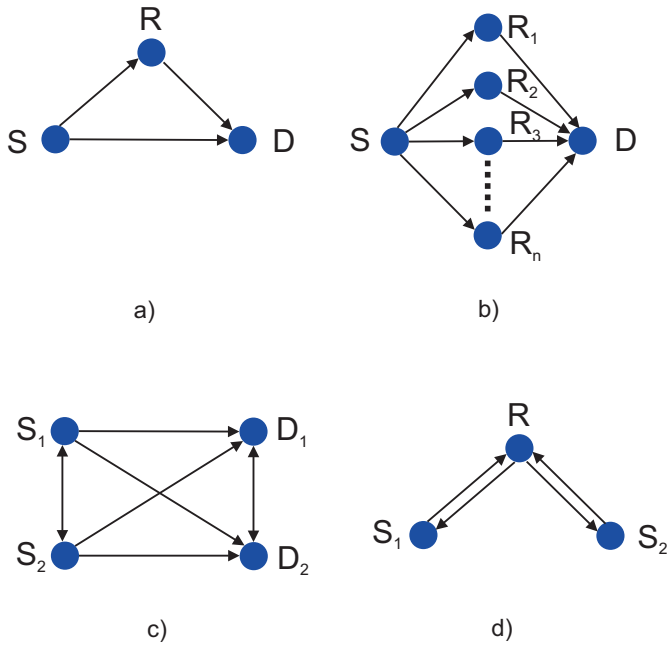


Figure 2.4: Different cooperation architectures: (a) classical three-node relay channel, (b) parallel relay channel, (c) virtual MIMO (V-MIMO) channel, (d) two-way relay channel network.

2.2.2 Relaying Protocols (DF, AF, and CF)

In a cooperative system, the relay node assists the communication with the destination node by either amplifying-and-forwarding (AF), decoding-and-forwarding (DF) or compressing-and-forwarding (CF) the received signal. In the AF operation mode, the relay node amplifies and retransmits the received signal, subject to its maximum transmit power constraint [40, 41]. Meanwhile, the signal received from the source terminal is demodulated and decoded before retransmission if the DF mode is employed. Finally, if a CF operation mode is implemented, a quantised and compressed version of the received signal is forwarded to the destination terminal using a source coding method [41]. In the following, different transmission types for cooperative systems, the principles of these relaying protocols and their different achievable rates are briefly introduced.

For each of the aforementioned cooperation schemes (AF, DF and CF) and using as a model the three-node classical relay channel, three different time-division multiple-access (TDMA)-based transmission protocols can be considered. For the transmission protocol Type I, the source node communicates with the relay and destination node during the first time slot. In the second time slot, both the relay and source node communicate with the destination node [40]. During the operation Type II, only the relay node transmits to the destination node over the second time slot [42]. Finally, the third Type is identical to the Type I apart for the fact that there is no direct transmission from source to destination during the first time slot [43]. The TDMA transmission protocols are then summarised in Table 2.1 [40].

Time Slot / Type	I	II	III
1	S $\rightarrow$ R, D	S $\rightarrow$ R, D	S $\rightarrow$ R
2	S $\rightarrow$ D, R $\rightarrow$ D	R $\rightarrow$ D	S $\rightarrow$ D, R $\rightarrow$ D

Table 2.1: TDMA-based protocols. S, R and D stand for the source, relay and destination nodes, respectively. A$\rightarrow$ B represents a communication between nodes A and B.

In order to proceed with the introduction of the relaying protocols, first the three-node classical relay channel is presented, where the subscripts s, r, and d are used to classify the transmitted (or received) signals at the source, relay and destination, respectively. The signal transmitted by the source during the i-th time slot is denoted by $x_{s,i}$ for $i = 1, 2$, with $\mathbb{E}[x_{s,i}] = 0$ and $\mathbb{E}[|x_{s,i}|^2] = 1$. Thus, the signal received at the relay and destination nodes in the first time slot

are given by

$$y_r = \sqrt{P_s} h_{sr} x_{s,1} + z_{r,1} \tag{2.10}$$

$$y_{d,1} = \sqrt{P_s} h_{sd} x_{s,1} + z_{d,1} \tag{2.11}$$

where P_s is the transmit power of the source node over one symbol period. Furthermore, the data channels from the source to the node $j \in \{r, d\}$ are represented by $h_{sj} x_{s,1}$ where h_{sj} is the small scale fading gain modelled as Rayleigh fading. Moreover, $z_{j,i} \sim \mathcal{CN}(0, \sigma_N^2)$ represents the white noise over the i-th time slot at the j-th node. During the second time slot, the destination node receives a superposition of the relay and source transmission according to

$$y_{d,2} = \sqrt{P_s} h_{sd} x_{s,2} + \sqrt{P_r} h_{rd} x_r + z_{d,2}, \tag{2.12}$$

where P_r denotes the relay transmit power and x_r is the re-transmitted signal from the relay to the destination node over the second time slot.

2.2.2.1 Amplify-and-Forward

Under an amplify-and-forward protocol, the relay receives a noisy version of the signal transmitted by the source than then amplifies and retransmits to the destination node. As shown in Table 2.1, in the first time slot, a data symbol $x_{s,1}$ is transmitted by the source to the destination. Over the second time slot, the destination receives a superposition of the relay transmission and the source transmission. Thus the received signal at the destination over the two time slots can be written as follows

$$y_{d,1} = \sqrt{P_s} h_{sd} x_{s,1} + z_{d,1}, \tag{2.13}$$

$$y_{d,2} = \sqrt{P_s} h_{sd} x_{s,2} + \sqrt{P_r} h_{rd} h_{sd} a x_{s,1} + \hat{z}_{d,2}, \tag{2.14}$$

where

$$a = \sqrt{\frac{E_r}{|h_{sr}|^2 E_{s,1} + \sigma_N^2}}, \tag{2.15}$$

is a normalization parameter, $E_{s,1} = P_s \mathbb{E}[|x_{s,1}|^2]$ is the source power during the first time slot, $E_r = P_r \mathbb{E}[|x_r|^2]$ is the relay power with $x_r = a y_r$ and the accumulative noise term is given by

$\hat{z}_{d,2} \sim \mathcal{N}(0, \sigma_N^2 + \sigma_N^2 |h_{rd}|^2 a^2)$. Once the data symbols have been detected, the achievable rate for the AF protocol, R_{AF}, can be obtained as [24, 42]

$$\mathrm{R}_{AF} = \frac{1}{2} \log_2 \left(1 + |h_{sd}|^2 E_{s,1} + \frac{|h_{sr}|^2 |h_{rd}|^2 E_r}{\sigma_N^2 + |h_{sr}|^2 E_{s,1} + |h_{rd}|^2 E_r} \right). \tag{2.16}$$

2.2.2.2 Decode-and-Forward

During the first time slot of the decode-and-forward protocol, the received signal at the relay is given by (2.10). In the second time slot, the relay node demodulates, decodes, re-encodes and retransmits the signal $x_r = \hat{x}_{s,1}$, where $\hat{x}_{s,1}$ denotes the estimated signal of $x_{s,1}$. Assuming that $x_{s,1}$ is successfully decoded at the relay, then the received signals at the destination over the two time slots are given by

$$y_{d,1} = \sqrt{P_s} h_{sd} x_{s,1} + z_{d,1}, \tag{2.17}$$

$$y_{d,2} = \sqrt{P_s} h_{sd} x_{s,2} + \sqrt{P_r} h_{rd} x_{s,1} + \hat{z}_{d,2}, \tag{2.18}$$

From [44] it is known that the achievable rate for the DF relaying protocol can be obtained as

$$\mathrm{R}_{DF} = \min\{\mathrm{R}_{DF1}, \mathrm{R}_{DF2}\} \leq \min\{I(x_{s,1}; y_r, y_d | x_r), I(x_{s,1}; x_r | y_d)\}, \tag{2.19}$$

where $I(x_{s,1}; y_r, y_d | x_r)$ upper bounds the mutual information of the broadcast channel, i.e., the rate of information transferred from the source to the relay and destination nodes. Meanwhile, the mutual information of the multiple-access channel, i.e., the channel between the source and relay nodes and the destination, is bounded by $I(x_{s,1}; x_r | y_d)$. As a result, the terms R_{DF1} and R_{DF2} in (2.19) are given by [13, 45]

$$\mathrm{R}_{DF1} = I(x_{s,1}; y_r, y_d | x_r) = \frac{1}{2} \log_2 \left(1 + |h_{sr}|^2 E_{s,1} \right), \tag{2.20}$$

$$\mathrm{R}_{DF2} = I(x_{s,1}; x_r | y_d) = \frac{1}{2} \log_2 \left(1 + |h_{rd}|^2 E_r + |h_{sd}|^2 E_{s,2} \right). \tag{2.21}$$

Substituting equations (2.20) and (2.21) into (2.19), the achievable rate for the DF relaying protocol can be written as

$$\mathrm{R}_{DF} = \frac{1}{2} \min \left\{ \log_2 \left(1 + |h_{sr}|^2 E_{s,1} \right), \log_2 \left(1 + |h_{rd}|^2 E_r + |h_{sd}|^2 E_{s,2} \right) \right\}. \tag{2.22}$$

2.2.2.3 Compress-and-Forward

During the compress-and-forward protocol, the relay forwards a quantised and compressed version of the received signal by employing standard quantisation, or some source coding techniques to compress the signal. After the second time slot and according to [46], the three-node relay system is equivalent to a system in which destination node has two antennas and thus the corresponding received signals are given by

$$\begin{bmatrix} y_r + z_c \\ y_{d,1} \end{bmatrix} = \begin{bmatrix} \sqrt{P_s} h_{sr} x_{s,1} + z_{r,1} + z_c \\ \sqrt{P_s} h_{sd} x_{s,1} + z_{d,1} \end{bmatrix}, \tag{2.23}$$

where z_c is the independent and identically distributed (i.i.d.) circular Gaussian noise, frequently referred to as compression noise, which is independent of y_r and $y_{d,1}$. Besides the widely used Wyner-Ziv (WZ) coding, a standard source coding can be employed for practical considerations. These include a reduced complexity at a comparable achievable rate [47] and the fact that there is no improvement in performance for non-correlated systems [48]. As a result, the signal received at the destination node during the second time slot is given by

$$y_{d,2} = \sqrt{P_s} h_{sd} x_{s,2} + \sqrt{P_r} h_{rd} x_{s,1} + z_{d,2} + z_c. \tag{2.24}$$

Assuming a standard source coding, then the power level of the compression noise z_c is lower bounded by [47]

$$\sigma_{CF}^2 = \frac{\mathbb{E}[||y_{d,2}||^2]}{2^C - 1}, \tag{2.25}$$

where C is the capacity of the error-free conference link between the relay and the destination nodes and is assumed to be equal to the fixed quantisation rate. Thus, the achievable rate of the CF protocol can be expressed as

$$\mathrm{R}_{CF} = \frac{1}{2} \log_2 \left(1 + |h_{sd}|^2 E_{s,1} + \frac{|h_{sr}|^2 E_{s,1}}{1 + \sigma_{CF}^2 + \sigma_{CN}^2} \right). \tag{2.26}$$

2.2.3 Virtual MIMO Systems

Cooperative systems are considered as an effective approach to improve spectral and energy efficiency of wireless networks without requiring multiple number of antennas per device. The

basic principle of a cooperative wireless system is to group an arbitrary number of single-antenna devices into virtual antenna arrays to mimic and exploit the gains of multi-antenna systems by means of wireless cooperation [49–51]. Virtual-MIMO (V-MIMO) systems are considered a particular relaying configuration (see Figure 2.4(c)), that uses distributed antennas on multiple devices to achieve some of the benefits provided by conventional MIMO Systems. In this section, the application of the classic three-node relay network is extended to a system with multiple source, relay and destination nodes, as shown in Figure 2.5.

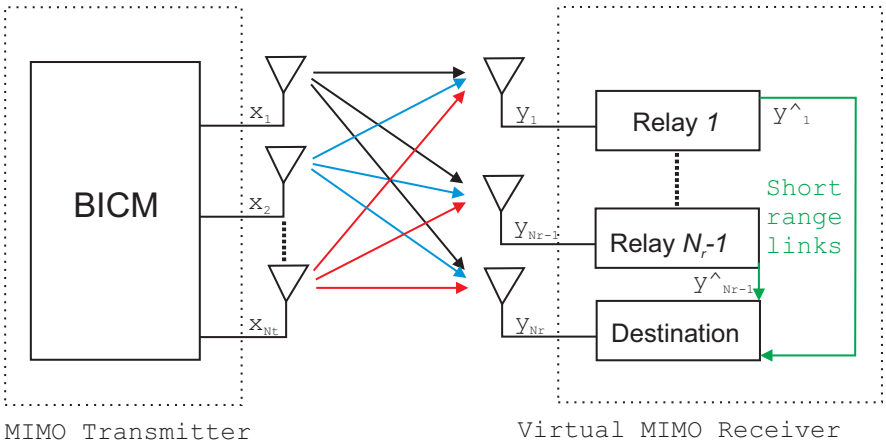


Figure 2.5: System model of the cooperative virtual MIMO system, that assumes a multi-antenna transmitter and single-antenna relay and destination nodes.

V-MIMO systems lie in the integration of MIMO and relay technologies. From the MIMO perspective, V-MIMO configurations are designed to achieve diversity, multiplexing or beam-forming gains. Meanwhile, from the relay perspective, these systems can adopt either AF, DF, CF or even a combination of these schemes to complete the transmission in one-hop or multiple hops [52]. Even though cooperative MIMO systems show an increase of the system complexity and a larger signalling overhead required for supporting the wireless cooperation, a cost-effective trade-off regarding capacity, cell edge throughput, coverage, mobility and complexity is obtained. These improvements hinge on the usage of distributed antennas, which increases the system capacity by means of spatial diversity [49]. When the separation between the relay and destination nodes varies, the AF, DF and CF protocols exhibit different performance in terms of capacity. If the relay is close to the destination, the CF protocol performs the better compared with DF [53, 54]. Considering that the power consumed for each of the protocols is relatively small compared with that of the RF circuitry, the slightly different power consumption between the protocols can be neglected. As can be seen from Figure 2.6, when the relay channel is weak ($g < 1$), DF fails to take advantage of the cooperative system, since

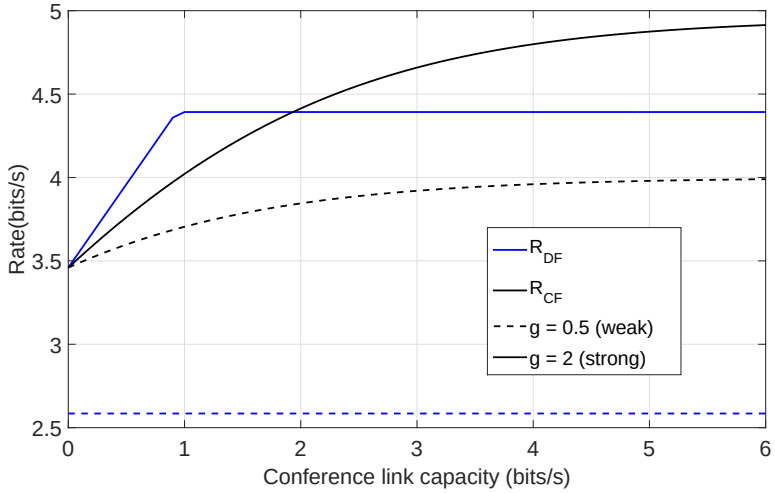


Figure 2.6: Achievable rates when a weak ($g = 0.5$) and strong ($g = 2$) relay channel is considered, where g is the relative channel strength of the relay and receiver [1].

its performance is limited by the rate at which the relay can correctly decode the received signal. On the other hand, when a reliable channel is considered ($g \geq 1$), DF outperforms CF and meets the cut-set bound of the capacity (C) of the conference link [1]. In other words, for large values of g (i.e., the relative channel strength of the relay and receiver), the DF not only outperforms the CF protocol but also is capacity-achieving (see Figure 2.7).

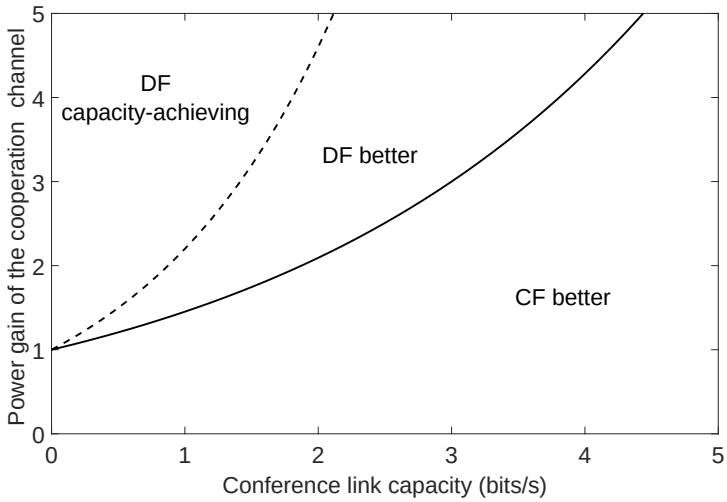
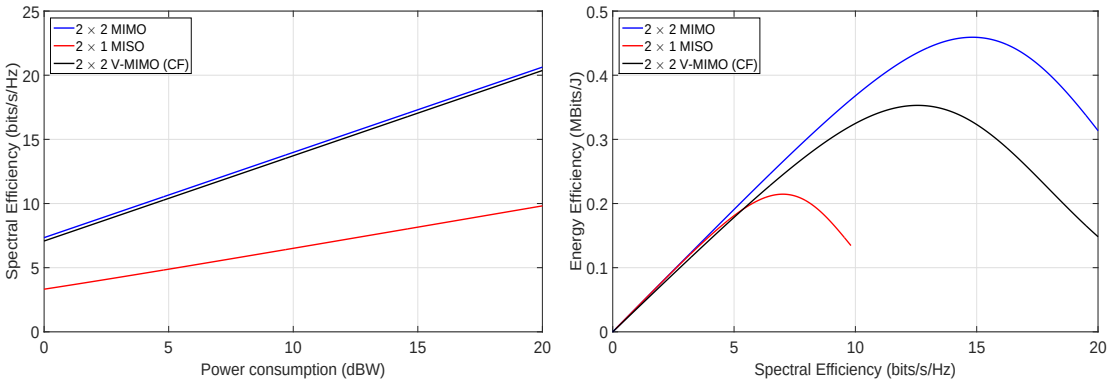


Figure 2.7: Best cooperation protocol based on the capacity and power gain of the conference link [1].

For illustrative purposes, Figures 2.8(a) and 2.8(b) show, respectively, the spectral efficiency (SE) and energy efficiency (EE) performance of a 2×1 MISO system, 2×2 MIMO system and 2×2 V-MIMO system. The physical channel propagation parameters are adopted from

the 3GPP LTE standard models [55] and equations (2.3) and (2.5) are used to calculate the SE for the MIMO and MISO channels, respectively. Meanwhile, the EE, defined as the number of bits that can be successfully delivered per unit of energy, was computed using the power consumption model introduced in Section 4.3. The relay implements a CF protocol through an out-of-band short-range cooperation channel and it is assumed that the maximum transmit power of the relay is smaller than that of the source terminal. From Figure 2.8(a), it can be noticed that the SE achieved by the cooperative MIMO system is comparable with that of a traditional MIMO system and much larger of that of a MISO system. Moreover, from Figure 2.8(b), it is possible to notice that in the low-SE regime, V-MIMO shows a much better EE performance than the non-cooperative MISO system and performs close to the ideal MIMO. Meanwhile, in the high-SE regime, V-MIMO performs much better than MISO but it is outperformed by the MIMO system, since the additional power required to carry out the wireless cooperation reduces its EE performance.



(a) (b)

Figure 2.8: Comparison of the (a) spectral and (b) energy efficiency performance for the V-MIMO that implements CF wireless cooperation, MISO and MIMO systems.

In this chapter, only the optimal vector quantiser (VQ), i.e., the Voronoi VQ (see Section 3.4.1.1 for further details) was considered. Nevertheless, it is evident that the practicality of CF cooperation will be greatly enhanced if an efficient source coding technique can be used at the relay. Thus, low complexity quantisers that enable the use of a system with large number of antennas will be presented in Chapter 3.

2.3 Millimetre Wave Wireless Systems

Despite research efforts to deploy efficient wireless technologies, mobile data traffic demand has experienced a unprecedented growth [56]. Recent studies suggest that millimetre-wave (mm-wave) frequencies could be used as an alternative to the currently saturated 700 MHz to 2.6 GHz radio spectrum bands for wireless communications [2]. In this section, the concept of mm-wave communications, its opportunities and challenges are introduced.

2.3.1 Mm-wave spectrum

Almost all commercial radio communications services operate in a narrow band of the radio frequency (RF) spectrum between 300 MHz and 3 GHz due to its favourable propagation characteristics. Thus, the portion of the RF spectrum above 3 GHz has been largely unexploited for commercial wireless applications [2, 57]. Within the 3 - 300 GHz spectrum, up to 252 GHz can potentially be suitable for mobile wireless communications as shown in Figure 2.9. Since, the transmitted signals can experience an attenuation of about 15 dB/km at around 60 GHz, and up to tens of dBs in the range of 164 - 200 GHz [58] due to the oxygen and water vapour absorption, these bands are frequently excluded for mobile broadband applications as the transmission range in these bands will be limited.

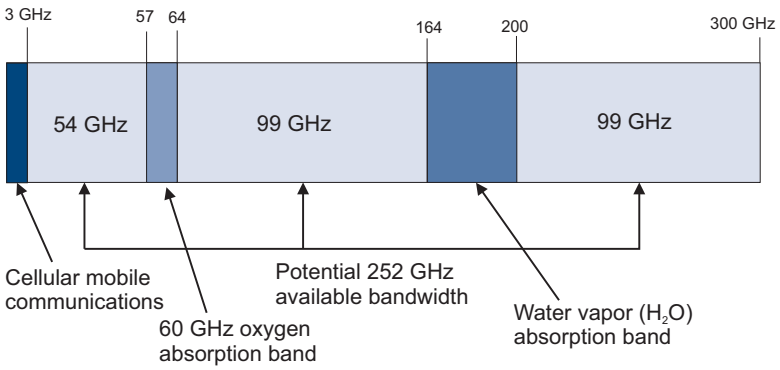


Figure 2.9: Millimetre wave spectrum [2].

However, considering that current cell sizes in urban environments are on the order of 200 m, atmospheric absorption does not create significant additional path loss for mm-wave frequencies. Only 7 dB/km of attenuation is expected during heavy rainfall, which can be translated to only 1.4 dB of attenuation over the 200 m of coverage [59, 60].

2.3.2 Millimetre Wave Channel Model

As mentioned in Section 2.1.4, Massive MIMO technology and small-cells are two promising approaches for future cellular networks. While, Massive MIMO can accurately concentrate transmitted energy to the mobile users, small cells decrease the average distance between transmitters and users by overlaying small cell access points, resulting in lower propagation losses and higher energy efficiency [61, 62]. Both of these approaches are readily supported and enhanced by the mm-wave spectrum, since the small wavelengths not only allow for hundreds of antenna elements to be placed in relatively small physical space, but also because by reducing the serving distance, mm-wave frequencies can overcome attenuations due the propagation. Furthermore, beamforming techniques can be employed to increase the link capacity in the mm-wave regime. Beamforming is essentially a spatial filtering operation typically using an array of antennas to radiate energy in a specific direction [63]. The gain obtained through antenna beamforming can also compensate for the path loss, improve SNR, enhance the Ricean factor gain, and reduce delay spread due to multipath dispersion [63, 64].

In order to characterise the channel for mm-wave communications detailed statistical models have been developed [58, 59, 65, 66]. Since range is one of the key issues that mm-wave frequencies face then it is important to determine how path loss varies with distance. In standard urban cellular models, it is common to fit the line-of-sight (LoS) and nonline-of-sight (NLoS) path losses separately. While LoS path losses roughly follow the free space propagation based on Friis' law, in order to characterise the presence of NLoS, frequently a standard linear fitting is used as follows [2, 5, 57]

$$PL(dB) = a + 10b\log_{10}(r) + \xi, \tag{2.27}$$

where $\xi \sim \mathcal{N}(0, \sigma_N^2)$ with σ_N^2 as the log-normal shadowing variance, r is the distance in meters, a and b are the least square fits. The values of a, b and σ_N are shown in Table 2.2.

Variable	28 GHz	73 GHz
NLoS	$a = 72$, $b = 2.92$, $\sigma_N = 8.7$ dB	$a = 82.7$, $b = 2.69$, $\sigma_N = 7.7$ dB
LoS	$a = 61.4$, $b = 2.0$, $\sigma_N = 5.8$ dB	$a = 69.8$, $b = 2.0$, $\sigma_N = 5.8$ dB

Table 2.2: Statistical model for the large-scale parameters [5].

All these parameters represent large-scale fading characteristics, thus, they are associated with

the macro-scattering environment and change relatively slowly [67]. In order to generate a random narrowband time-varying channel, the angles of arrival (AoA) and departure (AoD) can be used to describe each of the subpaths. Assuming a mm-wave system with N_r-receive antennas and N_t-transmit antennas, the narrowband time-varying channel gain between a transmitter-receiver pair can be represented by a matrix given by [5]

$$\mathbf{H} = \sqrt{\frac{1}{N_l}} \sum_{k=1}^{K} \sum_{l=1}^{N_l} \psi_{k,l} \mathbf{a}_r(\phi_{k,l}^r, \theta_{k,l}^r) \mathbf{a}_t^H(\phi_{k,l}^t, \theta_{k,l}^t), \tag{2.28}$$

where $\psi_{k,l}$ is the complex small-scale fading gain on the l-th subpath of the k-th cluster, $\mathbf{a}_r(\cdot)$ and $\mathbf{a}_t(\cdot)$ are the vector response functions for the receive and transmit antenna arrays to the angular arrivals and departures and $\phi_{k,l}^j$ and $\theta_{k,l}^j$ are the horizontal and vertical angles of arrival ($j = r$) or departure ($j = t$), respectively. Further details about the time-varying channel can be found in Section 5.3.1.

2.3.3 Challenges and Opportunities

Despite the potential of mm-wave communications, there are a number of key challenges to overcome in order to fully exploit the benefit of using mm-wave frequencies. Among these challenges, it is possible to point out:

- *Design of circuit components*: with high carrier frequency and wide bandwidth in mm-wave communications, several technical challenges in the design of circuit components and antennas need to be addressed. From severe non-linear distortion of power amplifiers (PA) to phase noise and IQ imbalance faced by the RF integrated circuits [68].
- *Interference Management*: since practical mm-wave systems are limited in range a large number of mm-wave access points may be deployed to provide coverage, thus, interference management mechanisms to prevent degradation of network performance could be required [69].
- *Blockage*: in order to ensure robust network connectivity due to the propagation characteristics of mm-wave frequencies, different approaches from the physical layer to the network layer can be considered: exploiting reflection, relaying techniques, etc [57, 69].

Due to the inherent limitations of mm-wave propagation, mm-wave cellular systems require to

coexist with other systems, such as long-term evolution (LTE) and WiFi, in order to provide uniform, robust high capacity across different deployments [57, 70]. Thus, it is expected that mm-wave networks will be inherently heterogeneous. Therefore, in heterogeneous networks, or HetNets, short-rage smalls cells operating in the mm-wave frequencies will coexist with macrocells transmitting in the microwave frequencies. Interaction and cooperation between different kinds of networks become a key factor to solve the problems of mobility management, vertical handover and mobile data offloading in future wireless communications [57].

2.4 Stochastic Geometry for Cellular Networks

While wireless communications systems have been studied for decades, until just a few years ago, mathematical performance analysis of these networks was not possible without extreme simplifications. In this section, basic concepts from stochastic geometry and Poisson point processes are introduced. For further details, the reader is advised to consider [15, 71–74].

2.4.1 Point Process

Stochastic geometry is a branch of applied probability which allows the study of random phenomena on the plane or in higher dimensions. It is intrinsically related to the theory of point processes [75]. Nowadays, it is widely used in the context of communication networks [76]. The most elemental object studied in stochastic geometry are the point processes. A *point process* can be depicted as a random collection of points in space (see Figure 2.10).

Formally, a point process (PP), $\Phi = \{\mathbf{X}_i, i \in \mathbb{N}\}$, is a random collection of points within a measure space. In the case of cellular networks this space is the Euclidean space $\mathbb{R}^d$ [72]. A convenient interpretation is to think of Φ as a counting measure, where the idea is to count the number of points falling in any set $A \subset \mathbb{R}^d$. Thus, a point process can be defined as [71]

$$\Psi(A) = \sum_{\mathbf{X}_i \in \Phi} \mathbb{1}(\mathbf{X}_i \in A), \tag{2.29}$$

where $\Psi(A)$ is a random variable whose distribution depends upon Φ. Since $\Psi(A)$ represents the number of points of Φ in A, then $\Phi(A) < \infty$ for any bounded $A \subset \mathbb{R}^d$. Recall that a set is bounded if it is contained in a ball with finite radius [73]. Generally speaking, point processes

in $\mathbb{R}^d$ exhibit some dichotomies as follows [76]:

- A PP can be either *stationary* or not. A PP is stationary if the law of the point process is invariant by translation. For example, if $\Phi = \{\mathbf{X}_i, i \in 1 \dots k\}$ is a PP, where $\mathbf{X}_i \in \mathbb{R}^2$ represents the location of each user of a cellular network. Then the distribution of the shifted PP $\Phi + v$ (obtained by shifting each point $\mathbf{X}_i \in \Phi$ to $\mathbf{X}_i + v$) is identical to the distribution of Φ.

- A PP can be *marked* or not. Marks assign labels to the points of the process to define some underlying property. They are typically independent of the PP and independent and identically distributed. For example, a PP formed by the location of the base stations is marked if each point is labelled as $\mathbf{X}_i = 0$ or $\mathbf{X}_i = 1$ according to their activity level.

- A PP can be *isotropic* or not. Isotropy holds if the law of the PP is invariant to rotation.

- A PP can be *Poisson* or not. A formal definition of the Poisson point process is given in the following subsection.

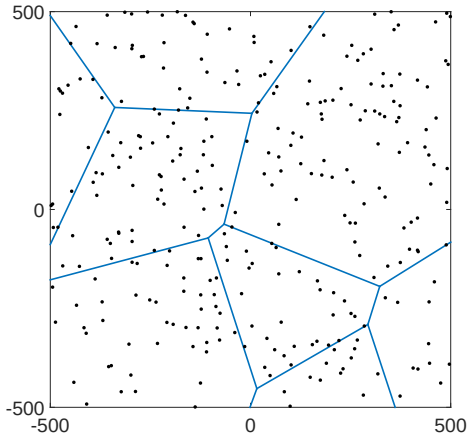


Figure 2.10: A single realization of a Point Process where the blue lines represent the Voronoi boundaries.

2.4.2 Poisson Point Process

A *Poisson point process* (PPP) is a point process with expected value $\mu(A) = \mathbb{E}[\Psi(A)]$ if [71, 72]

1. For all set A, the random variable $\Psi(A)$ is Poisson distributed with mean $\mu(A)$.

2. For any m disjoints sets $A_1, \ldots, A_m$, the random variables $\Psi(A_i)$ are independent.

Due to mathematically tractability, considerable research in cellular networks has been focused on the so called *homogeneous Poisson point process*, which essentially is a PPP with uniform intensity λ such that

$$\mu(A) = \lambda \ell(A), \tag{2.30}$$

where $\ell(A)$ is the Lebesgue measure (or size) of A. Some of the most useful properties of the PPP are listed below [71, 72]:

- *Independent thinning* of a PPP results in a different PPP. If each point in a PPP is randomly and independently assigned binary $\{0, 1\}$ marks with $\mathbb{P}(Q_i = 1) = q$, then the collection of points marked as 1 form a new PPP with intensity $q\lambda$.
- *Superposition* of independent PPPs results in a PPP. Thus, if m independent homogeneous PPPs characterized by intensities λ_i with $i = 1, 2, \ldots, m$ are combined to form a new PPP, this new PP will also be a PPP, now with intensity $\sum_{i=1}^{m} \lambda_i$.
- *Displacement of a PPP* results in a different PPP. Thus, if each point of a PPP is displaced by some random law, for example by adding i.i.d Gaussian random variables to each point, the PP consisting of these new random points will also be PPP.

2.4.3 Downlink Analysis

Consider the downlink of a cellular network that consist of base stations (BSs) located according to a homogeneous PPP, $\Phi^{(BS)}$ of intensity λ in the Euclidean plane and the location of the users (UEs) independently determined by other stationary point process, $\Phi^{(UE)}$. It is assumed that each user is associated with the closest base station, i.e., the users in the Voronoi cell of a BS are associated with it (see Figure 2.10). As a result, the user is served by the BS that provides it the highest average SINR. Moreover, it is assumed that the transmission between the base station and the user is constant with a fixed power P following a orthogonal multiple access within a cell. Thus, the intended user sees interference from all other BSs in the network except from its tagged BS. Once the signal is transmitted, it attenuates with distance according to the standard power-law path loss propagation model, where $\alpha > 2$. Thus, the received signal at a distance r has an average power of $P(r) = Pr^{-\alpha}$. In order to take into account the channel effects, multiplicative random values H for the desired signal and I_i for the i-th interferer

are employed. For the sake of simplicity, it is assumed that these values follow a Rayleigh distribution with unit power.

In the following the SINR, servicing distance, interference, and probability of coverage of this cellular network are briefly introduced. Further details can be found in [15, 71–75, 77, 78] and the references therein.

2.4.3.1 Distance to the nearest Base Station

An important measure of the cellular network is the distance r separating a typical user from its tagged BS. From [71], it is known that no other BS can be closer that r, since each user is served by the closest BS. In other words, all interfering BSs are located farther than r. Using the fact that the null probability of a Poisson process within an area of interest A is $\exp(-\lambda A)$, then the probability (PDF) and cumulative (CDF) density function of r are given by [15]

$$f_R(r) = 2\pi r\lambda \exp(-\pi\lambda r^2), \qquad r \geq 0, \tag{2.31a}$$

$$F_R(r) = 1 - \exp(-\pi\lambda r^2), \qquad r \geq 0. \tag{2.31b}$$

2.4.3.2 Signal-interference-to-noise-ratio (SINR)

Without loss of generality and using the Slivnyak's theorem that states that conditioning a PPP on a point at x does not change the distribution of the rest of the process, it is assumed that the user under consideration is located at the origin. Then, the SINR of the user at a random distance r from its tagged BS can be expressed as

$$\mathit{SINR} = \frac{HPr^{-\alpha}}{\sigma_N^2 + I_R}, \tag{2.32}$$

where

$$I_R = \sum_{X_i \in \Phi \setminus b(0,0)} G_i P ||X_i||^{-\alpha}, \tag{2.33}$$

is the cumulative interference from all the other BSs, excluding the tagged BS for a typical user at the origin, which is denoted by $b(0,0)$. For simplicity the interference I_R is considered a

standard M/M shot noise [73, 74, 79], that is created by a Poisson point process of intensity λ outside a disc, $b(0, R)$, centred at the origin of radius R.

2.4.3.3 Probability of Coverage

The probability of coverage is the probability that a typical user is able to achieve some threshold SINR, τ, i.e., it is the complementary cumulative distribution function (CCDF) of SINR and is defined by [71, 73]

$$\mathbb{P}_C = \mathbb{P}(\textit{SINR} > \tau). \tag{2.34}$$

Equation (2.34) can also be thought as the average fraction of users who at any time achieve a target SINR τ. Conditioning on the nearest BS being at a distance r from the typical user, the probability of coverage of the homogeneous PPP network is given by [71]

$$\mathbb{P}_C = \int_{r>0} \mathbb{P}(\textit{SINR} > \tau | r) f_R(r) dr. \tag{2.35}$$

Using equation (2.31a), the fact that H is an exponential random variable with unit mean, i.e., $H \sim \exp(1)$ and some algebraic manipulation, the probability of coverage can be simplified to

$$\mathbb{P}_C = 2\pi\lambda \int_{r>0} e^{\pi\lambda r^2} e^{-\tau P^{-1} r^\alpha \sigma_N^2} \mathcal{L}_{I_R}\left(\tau P^{-1} r^\alpha\right) r dr, \tag{2.36}$$

where $\mathcal{L}_{I_R}\left(\tau P^{-1} r^\alpha\right)$ is the Laplace transform of equation (2.33), and is given by

$$\mathcal{L}_{I_R}\left(\tau P^{-1} r^\alpha\right) = \exp(\pi r^2 \lambda \rho(\tau, \alpha)), \tag{2.37}$$

where

$$\rho(\tau, \alpha) = \tau^{2/\alpha} \int_{\tau^{-2/\alpha}}^{\infty} \frac{1}{1 + u^{\alpha/2}} du. \tag{2.38}$$

2.4.3.4 Validation

Practical LTE cellular networks are usually deployed in a more strategic manner than just random independent placing, and for this reason a regular hexagonal grid has been used to model

current cells. The hexagonal grid model can be considered as an upper bound since a perfectly regular geometry is in fact optimal from a coverage point of view [80]. However, this approach is idealised, since imperfections relative to the landforms and deployment costs produce in practice irregular grids. Simulation studies account for these imperfections by perturbing the BSs locations with a 2D Gaussian or uniform random variables [81]. Since a practical LTE network lies somewhere between the two extremes of perfect regularity (hexagonal grid) and complete randomness (PPP), it is expected that its SINR coverage probability be bounded by these two extremes, as pointed by [80].

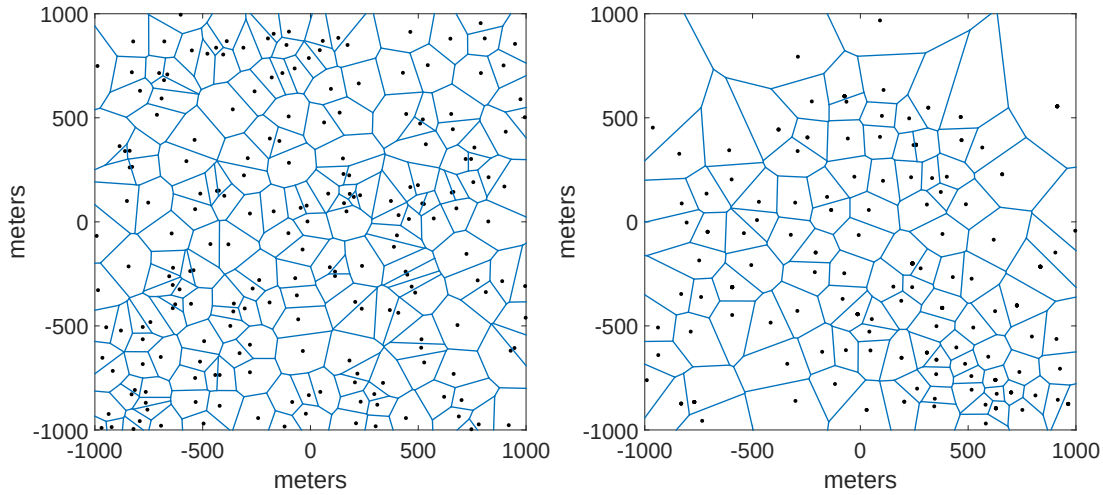


Figure 2.11: (a) Homogeneous PPP BSs deployment compared with a (b) 2000m $\times$2000m section of a LTE network in central London [3], where the black dots represent the BSs locations and the blue lines their corresponding coverage.

Figure 2.11 compares a 2000m $\times$2000m section of a LTE network in central London and a sample of BSs generated from a homogeneous PPP of the same density. The main weakness of the Poisson model is that due to the independence of the PPP, BSs will in some cases be located very close together with significant coverage area. This is balanced by two strengths: the characteristic inclusion of different cell sizes and shapes and the lack of edge effects, i.e., the network extends indefinitely in all directions [71]. Nevertheless, as it is shown in Chapter 5, despite the analytical tractability, the assumption of a homogeneous PPP is not always the most accurate model for future 5G wireless networks. Since the BSs are to be deployed in locations where the users are more likely to appear, having as result a network where the uses will be clustered towards the BS, rather than being uniformly distributed within the area of interest.

An illustrative plot is presented in Figure 2.12, where the probability of coverage of a regular

hexagonal grid and a homogeneous PPP is presented. From this plot, it is possible to notice that the curves have the same shape but there is a performance gap of about 4 dB over nearly the entire SINR range, which depends on the actual BS layout used and the path loss exponent. Thus, the probability of coverage of a practical cellular systems is expected to lie roughly between the grid and the PPP deployments.

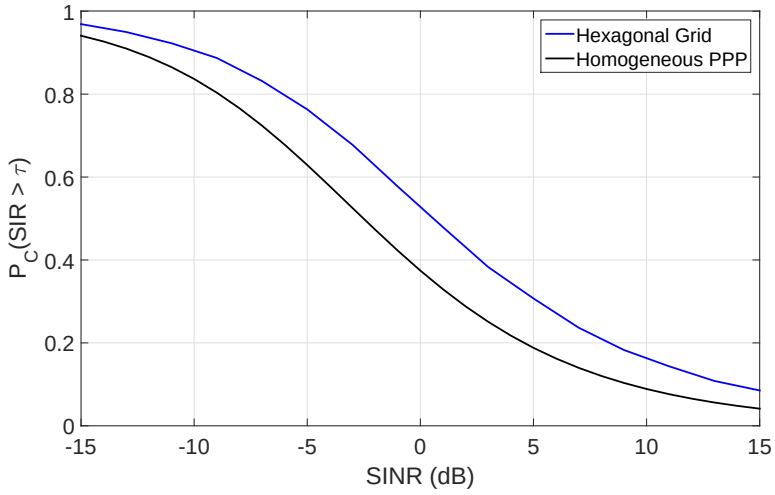


Figure 2.12: A comparison of the interference-limited coverage probability with $\alpha = 3$ for a 500m $\times$ 500m regular grid and a Poisson layout.

2.5 Summary

This chapter provided some basic principles on MIMO systems, cooperative communications, mm-wave systems and stochastic geometry. In the first section, the basic concepts for MIMO systems were introduced. It was shown that, MIMO systems can provide significant improvements over the conventional SISO case, in terms of both achievable rate and link reliability.

In the second section, cooperative wireless communication systems were introduced as a practical alternative to traditional MIMO systems, when the transmitter or receiver may not be able to support multiple antennas. Typical cooperation protocols, i.e., AF, DF and CF, and their achievable rates were illustrated. It was shown that, DF achieves a higher rate when the relay is close to the source, but CF outperforms DF when the relay gets closer to the destination.

In the third part, a brief review of mm-wave wireless systems was provided. The propagation characteristics of mm-wave frequencies were introduced and the opportunities and challenges were presented.

Finally, an introduction to stochastic geometry for modelling future 5G wireless communications systems was presented. Where the SINR, the serving distance and the probability of coverage of homogeneous PPP wireless systems were characterised.

Chapter 3
Low Complexity Codebook Design

In this chapter, a cooperative virtual MIMO (V-MIMO) system that uses a large multi-antenna transmitter and implements bit-interleaved coded modulation (BICM) transmission and compress-and-forward (CF) relay cooperation among multiple receiving nodes is presented. Since constructing a reliable codebook is the most computationally complex task performed by the relay nodes in CF cooperation, reduced complexity quantisation techniques are introduced in this chapter. In order to focus our analysis on the block error probability (BLER) and computational complexity, and considering that the relays and the destination nodes are closely spaced, an error-free conference link is assumed between them.

Two different scalar quantisation techniques are presented in this chapter, where the real and imaginary parts of the received signals are quantised independently. First, the uniform scalar quantiser (U-SQ) is considered and the performance achieved when a large number of antennas is used both at the transmitter and receiver is evaluated. Second, the Lloyd-Max algorithm (LM-SQ) [82] is employed to reduce the error distortion by taking advantage of the distribution of the received signals. A comparison in terms of BLER and computational complexity for designing the codebook and forwarding a compressed version of the signal to the destination node are presented. In practice, besides the data symbols, additional control information is transmitted, thus, the overhead incurred by the quantisation techniques is computed by calculating the number of bits required to transmit the codebook to the destination node.

Numerical results show that the LM-SQ is simpler to design and can achieve a BLER performance comparable to the optimal vector quantiser (V-VQ). Furthermore, due to its low complexity, U-SQ could be consider particularly suitable for very large wireless systems.

The remainder of this chapter is organised as follows. Section 3.1 presents the background and motivation of this chapter. The system model is described in Section 3.2. The CF cooperation at the relays is presented in Section 3.3. A description of the low complexity quantisation techniques, codebook design and the corresponding complexity analysis are presented in Section 3.4. Simulation results are shown in Section 3.5. Finally, Section 3.6 concludes the chapter.

3.1 Introduction

MIMO systems have drawn significant interest in recent years, due to their ability to increase both capacity and reliability of modern wireless communications. Nevertheless, design constraints for mobile devices limit the number of usable antennas; further, closely spaced antennas entail the presence of highly correlated fading signals, limiting performance. V-MIMO systems have recently emerged as a promising technique due to their ability to mimic and exploit the gains of multi-antenna systems by means of wireless cooperation [83]. A CF cooperation provides a good trade-off between performance, complexity and data rate when the relays are located close by the destination. In a CF - based cooperation, constructing a reliable codebook and forwarding a compressed version of the received signals to the destination node for subsequent processing are the main tasks that each relay has to perform. The codebook generation is the most computationally complex stage in the design [84].

In [85], codebook designs of the V-VQ and the tree-structured vector quantisation (TSVQ) to enable CF cooperation are presented. Different from the V-VQ, a multi-stage TSVQ where the encoding task is divided into several stages is introduced. It was shown that the TSVQ is much simpler to design and can achieve performance comparable to the optimal V-VQ. Meanwhile, an achievable rate based quantisation scheme is presented in [86]. This scheme based on a scalar quantiser achieves better performance and higher compression efficiency in the AWGN scenario.

In [87], the authors combined the improved affinity propagation (IAP) with the conventional Linde-Buzo-Gray (LBG) algorithm to generate an effective algorithm called IAP-LBG. According to the experimental results, the proposed method not only improved its convergence abilities but also provided higher-quality codebooks than conventional LBG method does. A fast search method based on principal component analysis (PCA) is proposed in [88]. The PCA-LBG-based VQ algorithm outperformed other methods in terms of peak signal-to-noise ratio and number of codewords searched.

Furthermore, in [89] a modified LBG algorithm is proposed. It was shown that, compared with the conventional V-VQ, it reduces the execution time and number of iterations by shifting all the low-utility codewords near to the cells with high utility. In [90], a channel quantised physical-layer network coding (CQ-PNC) scheme based on VBLAST to achieve the full diversity was presented. Specifically, the relay uses QR decomposition to convert the received signals and es-

timate the Gaussian integer summation. Numerical simulations showed that CQ-PNC performs within a 2 dB gap from the theoretical bound.

A universal unitary space vector quantisation (USVQ) codebook design criterion was provided in [91], which can design optimum codebooks for various spatial correlated channels with arbitrary antenna configurations. Simulations showed that the capacity achieved is very close to that obtained through ideal precoding and outperformed that of the schemes using the traditional Grassmannian codebooks. Further, a random vector quantisation approach that generates the codevectors independently from a uniform distribution on the complex unit sphere, using a limited number of feedback bits was presented in [92]. Meanwhile, the authors in [93] adopted the generalized Lloyd algorithm (GLA) to optimize the codebook in terms of maximal average signal-to-noise ratio (SNR) for a distributed beamforming with limited feedback for a time varying amplify-and-forward cooperative network. Numerical results showed that the distributed beamforming with compressed feedback performs closely to the case with infinite feedback.

In [94], the authors designed and analysed the performance of transmit beamformers for MIMO systems, based on bandwidth-limited information, by casting the design of transmit beamforming as an equivalent sphere vector quantisation (SVQ) problem. As a result, the rate-distortion performance achieved by the generalized Lloyd's algorithm is lower-bounded. In order to optimize symmetric and asymmetric scalable multiple description quantisers, in [95], an extension of the Lloyd-Max algorithm was described. Experimental results confirmed that, compared to conventional schemes, the proposed quantiser account for a significant average gain in SNR for a wide range of packet loss rates.

Additionally, in [96], a vector quantisation technique that exploits the reduced dimensionality of the channel Gram matrix by using Jensen's inequality and eigen-decomposition to improve the pairwise error probability was investigated. Moreover, in [97], a initial codebook design method for VQ was proposed. The mean and variance characteristic values were used to partition the training vectors into different groups and the codewords were selected from each group to generate an initial codebook. Experimental results demonstrated that the proposed method exhibits a better performance than the conventional ones.

In this chapter, reduced complexity quantisation techniques are proposed for V-MIMO to enable a large cooperative system with multiple relays. This work differs from the previous research

in that i) it is proposed using scalar quantisation at each relay node, where the Lloyd - Max algorithm and uniform scalar quantisation are considered, ii) it analyses the complexity and iii) presents the block error rate performance of the proposed quantisation techniques. Simulation results show that LM-SQ is less complex and achieves a comparable performance to the optimal V-VQ. Meanwhile, U-SQ is much simpler and exhibits a good performance when a large V-MIMO array is considered, making it suitable for future wireless communications.

3.2 System Model

Consider a cooperative V-MIMO system with a N_t-antenna transmitter and a receiver comprising N_r single-antenna devices, as shown in Figure 3.1. At the transmitter side, a BICM scheme is employed to provide error correction and improve the system's performance. The data bits are encoded using a rate-R_b linear convolutional encoder and then passed through an ideal bitwise interleaver, which rearranges the coded bits using a random permutation. Next, the interleaved coded bits are demultiplexed into N_t streams. In each stream $k = 1, \ldots, N_t$ a group of m_s^b coded bits are mapped to a data symbol x_k from a symbol alphabet $\mathcal{X}$ of size $|\mathcal{X}| = 2^{m_s^b}$. Thus, the transmission rate is $N_t m_s^b R_b$ bits per channel use. Finally, the transmitter simultaneously sends the modulated data symbols through N_t antennas using a transmit vector given by $\mathbf{x} = [x_1, \ldots, x_{N_t}]^{\mathrm{T}} \in \mathcal{X}^{N_t}$ that satisfies the power constraint $\mathbb{E}[\| \mathbf{x} \|^2] = E_s/N_t$. Thus, the corresponding power per bit is $E_b = E_s/(m_s^b R_b)$.

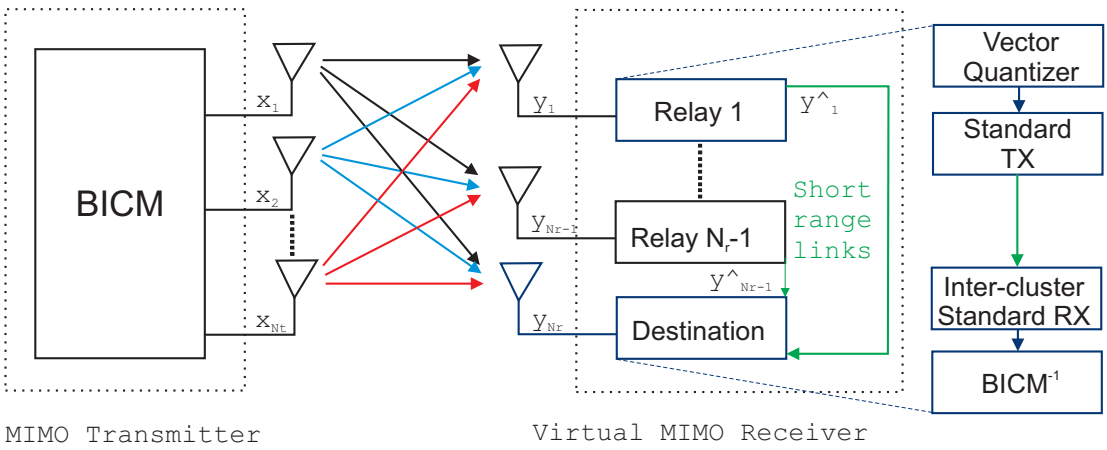


Figure 3.1: Block diagram of MIMO transmitter and virtual MIMO receiver.

In order to model the channel effects and considering that the transmitter is located far away from the receivers, a block fading channel is employed. Since, perfect channel state information

(CSI) is assumed at the receivers only, then, the received signal is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{H}$ denotes the $N_r \times N_t$ channel matrix and $\mathbf{n}$ the noise vector with components $n_l \sim \mathcal{CN}(0, N_0)$ and $l = 1, \ldots, N_r$. The notation $\mathcal{CN}(\mu, \sigma^2)$ represents a complex Gaussian distribution with mean μ and variance σ^2. Since, a slow fading propagation model is assumed, $\mathbf{H}$ remains constant through L symbol periods.

The cluster of receivers, that includes the destination node and the $N_c = N_r - 1$ relays, are assumed to be closely spaced, thus, it is reasonable to expect that the wireless cooperation between the relays and destination is considerably better and more stable than that between the transmitter and receivers. In fact, it is highly likely that a high channel capacity could be achieved due to the high reliability of these short range links. As was considered in [1, 24, 98, 99], it is assumed that the relays and destination nodes cooperate through error-free conference links of capacity C. The one-shot cooperation links require the destination to decode the signals sent over the conference links. In order to handle the amount of data that is to be sent from the relays to the destination, in this chapter, the CF protocol is considered. Hence, a compressed version of the signals will be passed to a standard transmitter and sent over the conference link, as will be detailed in Section 3.4.

Moreover, the destination node is assumed to be equipped with a traditional receiver which observes the signals from the transmitter, and an intra-cluster receiver which observes signals from the relays. Thus, the cooperation is performed at each relay through the simultaneous and independent transmission of orthogonal signals to minimise the inter channel interference. Later, the destination performs zero forcing (ZF) demodulation of the received signal and computes the log-likelihood ratio (LLR) for each coded bit. The multi-path LLRs are combined into one output stream by the multiplexer, and then rearranged into their original positions by the deinterleaver. Finally, the decoder accepts the LLRs of all coded bits and employs a soft-input Viterbi algorithm to decode the signals. Thus, with help from the relays, the single-antenna destination node receives N_r-path signals. Hence, a reliable compression process would allow the destination node to handle the compression noise and perform MIMO decoding [24].

3.3 Compress-and-Forward Cooperation

Aiming to mimic the performance of a traditional MIMO system, a V-MIMO operation is carried out at the receiver side, where N_c relays cooperate with the destination node through error-free conference links. To avoid interfering with the data sent by the transmitter, the conference links are assumed to be carried out via orthogonal channels to the transmitter array, i.e., different frequencies bands are used for transmitting and receiving signals at each relay. Compared with the long range data channel $\mathbf{H}$, the orthogonal conference links are shorter-range and could be reused many times within the coverage area. Consequently, the wireless cooperation will not affect the overall system throughput. In Chapter 2, different cooperative techniques including the AF, DF, and CF schemes were presented. Since the AF protocol amplifies not only its received signal but also the associated Gaussian noise, an AF-relayed packet may become too noisy to be decoded in practice. Moreover, the DF protocol does not take advantage of the aggregated information provided by the rest of the nodes, since it performs a single pass decoding at each relay and its computational complexity can be restrictive. Thus, a better candidate to implement the V-MIMO system is the CF protocol, as it provides a better performance when the relays are closer to the destination [46] and reduces the amount of data that needs to be sent through the short-range links.

Assuming that the CF protocol is used for wireless cooperation, then the main tasks performed at each relay includes designing a codebook, quantising the received signals and forwarding the compressed signals. After a quantiser is employed at each relay, a compressed version of the signal will be passed through the conference links. In [99] and [100], it was demonstrated that if the compressed version of the signal at each relay, $\hat{y}_c$, follows a Gaussian distribution, then the compression noise can be modelled by an i.i.d complex Gaussian variable as follows

$$\hat{y}_c = y_r + \hat{n}_c \tag{3.2}$$

where y_r is the received signal at each relay and $\hat{n}_c \sim \mathcal{CN}(0, \sigma^2_{CF})$ with $c \in \{1, \ldots, N_c\}$ is the compression noise [46, 84], which is assumed to be independent of y_r and can be computed according to equation (2.25) as $\sigma^2_{CF} = \frac{\mathbb{E}[||y_c||^2]}{2^C - 1}$. Along with the compressed version of the received signal, each relay node passes to the destination node information related with σ^2_{CF}. Once the destination knows the value of σ^2_{CF}, it scales the received signals using a degradation factor η_c per stream so that the relayed signals and the one received directly from the transmitter

have the same power of additive Gaussian noise [101, 102]. Thus, the cooperative system can be modelled as follows

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{H}}\mathbf{x} + \mathbf{n}, \tag{3.3}$$

where

$$\widetilde{\mathbf{H}} \triangleq \begin{bmatrix} \sqrt{\eta_1}h_{1,1} & \cdots & \sqrt{\eta_1}h_{1,N_t} \\ \vdots & \ddots & \vdots \\ \sqrt{\eta_{N_c}}h_{N_c,1} & \cdots & \sqrt{\eta_{N_c}}h_{N_c,N_t} \\ h_{N_r,1} & \cdots & h_{N_r,N_t} \end{bmatrix} \text{ and } \eta_c \triangleq \frac{N_0}{N_0 + \sigma_{CF}^2}. \tag{3.4}$$

Notice that the r-th row of $\widetilde{\mathbf{H}}$ corresponds to the direct link between transmitter and destination node. Since there is no relay assisting the destination node then there is no degradation factor affecting the transmission and thus $\eta_r = 1$. The receiver implements a soft demodulator with low complexity by using a linear ZF equalizer followed by per-layer max-log LLR calculation. Following the model presented in [103], the LLR for each coded bit c_k^λ is calculated by

$$\mathcal{L}(c_k^\lambda|\widetilde{\mathbf{y}}, \widetilde{\mathbf{H}}) = \frac{1}{\sigma_k^2}\left[\min_{x\in\mathcal{X}_i^0} |\hat{x}_k - x|^2 - \min_{x\in\mathcal{X}_i^1} |\hat{x}_k - x|^2\right], \tag{3.5}$$

for $i = 1, \ldots, \log_2(M)$ and $\lambda \in \{1, \ldots, m_s^b \times N_t\}$, where $\mathcal{X}_i^b$ denotes the set of symbols in the M-QAM constellation whose bit label at position i equals $b \in \{0, 1\}$, $\hat{x}_k$ is an estimate of the symbol in layer k provided by the equalizer, and σ_k^2 is an equalizer weight. In the ZF based demodulator, the equalization consists of

$$\hat{\mathbf{x}}_{\text{ZF}} = (\widetilde{\mathbf{H}}^{\mathbf{H}}\widetilde{\mathbf{H}})^{-1}\widetilde{\mathbf{H}}^{\mathbf{H}}\widetilde{\mathbf{y}} = \mathbf{x} + \widetilde{\mathbf{n}}, \tag{3.6}$$

where the post-equalization noise vector, $\widetilde{\mathbf{n}}$, has a covariance matrix given by

$$\mathbf{R} = \mathbb{E}\{\widetilde{\mathbf{n}}\widetilde{\mathbf{n}}^{\mathbf{H}}\} = N_0(\widetilde{\mathbf{H}}^{\mathbf{H}}\widetilde{\mathbf{H}})^{-1}. \tag{3.7}$$

Therefore, the approximate bit LLRs are obtained with the symbol estimate $\hat{x}_k = (\hat{\mathbf{x}}_{\text{ZF}})_k$ and weight factor $\sigma_k^2 = (\mathbf{R})_{k,k}$. The notation $(\mathbf{a})_k$ and $(\mathbf{A})_{l,k}$ represent the k-th element of vector $\mathbf{a}$ and the element in row k and column l of matrix $\mathbf{A}$, respectively. Calculating the LLRs separately for each layer results in a significant complexity reduction. In particular, the

complexity of evaluating (3.7) for all coded bits scales as $\mathcal{O}(N_t N_r |\mathcal{X}_i^b|)$, i.e., linearly with the number of antennas [103, 104].

3.4 Quantisation Design at the Relays

When the CF protocol is used for wireless cooperation, the main tasks performed at each relay node are: (a) designing a codebook, (b) quantising the received signals and (c) forwarding the compressed signals. Thus, the design of an efficient codebook is a critical factor to perform reliable CF cooperation. In this section, a standard source coding technique of rate R_c and measured in bits per/sample is employed. Moreover, it is assumed that the source coding rate R_c is equal to the error-free conference link capacity $2C$, where C is measured in bits per dimension/sample. Frequently, it is assumed that the relay nodes are equipped with a vector quantiser (VQ) module that performs an exhaustive search to find the correct codeword. Nevertheless, its high computational complexity make its implementation impractical for high-order modulations, large C or when a large-antenna system is considered. In this section, the widely used Voronoi vector quantiser (V-VQ) algorithm is presented and alternative algorithms for low complexity codebook design are introduced.

3.4.1 Codebook Design

During the codebook design, knowledge of the noise-free constellation and the size of the codebook 2^{2C} is required at the relay. Note that, in practice besides the data symbols and the modulation order, additional control information is transmitted through the control channel. Thus, the c-th relay is able to construct the noise-free constellation $y_c^{\mathrm{NF}} = \{h_c x_c + \sum_{i \in [1,\ldots,N_c] \setminus j} h_i x_i\}$ of its received signals.

3.4.1.1 Voronoi Vector Quantiser

A special class of vector quantisers are the so called *Voronoi* or *nearest neighbour* vector quantisers that has the feature that the partitions regions are completely determined by the codebook and distortion measure [105, 106]. In [107], Linde, Buzo, and Gray (LBG) proposed a VQ design algorithm based on a training sequence that bypasses the need for multi-dimensional integration and provides an optimal codebook in the sense of minimising the average distortion.

Specifically, if a mean squared error (MSE) distortion measure is assumed, then the encoding regions $\mathcal{D}_q$, are defined by the nearest neighbour criteria given by

$$\mathcal{D}_q = \{y_q : ||y_q - v_q||^2 \leq ||y_q - v'_q||^2, \forall q' = 1, 2, \ldots, 2^{2C}\}, \tag{3.8}$$

where $\mathcal{D}_q$ is the encoding region associated with the codevector v_q, y_q denotes the noise-free constellation of the received signals and the number of codevectors within the codebook is 2^C. This condition guarantees that the encoding region consists of all vectors that are closer to v_q than any of the other codevectors. Moreover, each codevector v_q should be averaged over the number of the training vectors within the encoding region as follows [107]

$$v_q = \frac{\sum_{y_q \in \mathcal{D}_q} v_q}{\sum_{y_q \in \mathcal{D}_q} 1}, \qquad q = 1, 2, \ldots, 2^{2C}. \tag{3.9}$$

In implementation, it is required that at least one training vector belongs to each encoding region to guarantee that (3.9) is finite. In general, given the number of codevectors required for quantisation 2^{2C}, the LBG is an iterative algorithm which provides a codebook in two steps. Using (3.8), it provides a set of partition regions (the set of black lines) and, through (3.9), updates the codevectors (the set of all red stars), which result in the smallest MSE distortion. Figure 3.2 shows the partition regions and codebook obtained when the V-VQ is employed. In practice, every received signal falling in a particular region are approximated by the red star associated with that region. Note that in this plot there are 32 regions and 32 codevectors, each of which can be uniquely represented by 5 bits. Thus, this is a 2-dimensional, 5-bit V-VQ.

3.4.1.2 Uniform Scalar Quantiser

Unlike vector quantisation that maps a n-dimensional vector per unit time, a quantisation scheme is called scalar if the source signal is quantised one sample at a time [108]. Although this quantiser is not ideal in terms of minimising the distortion, scalar quantisation is a rather simple technique that can be easily implemented in hardware. Scalar quantisers can be classified as uniform or non-uniform. While the decision boundaries in uniform quantisation are equally spaced, the non-uniform schemes try to assign the size of decision intervals to be inversely proportional to the probability density [105, 108]. In uniform scalar quantisation each codeword can be represented using $C = \lceil \log_2 N_i \rceil$ bits per dimension, where N_i is the number

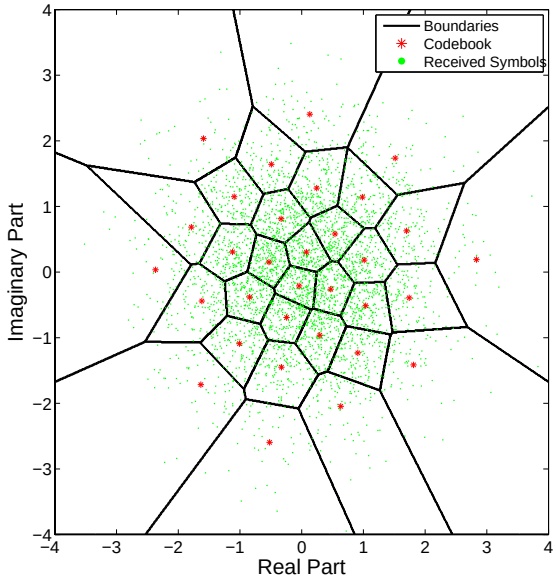


Figure 3.2: Two-dimensional Voronoi vector quantiser based on the LBG algorithm with $R_c = 5$ bits/sample. The solid black lines define the partition regions and the red starts denote the codevectors.

of decision intervals and C is referred to as the bits per dimension/sample. Since the contribution of each error to the MSE is weighted by the probability density function (PDF) of the signal [108], the choice of the step size, or equivalently the range $[x_{\min}, x_{\max}]$ determines the quantised error obtained. Recall that the MSE not only accounts for the granular error but also the overload distortion. As shown in Figure 3.3, if the source signal is not bounded then the dynamic range of a uniform quantiser cannot cover the whole range. The areas beyond $[x_{\min}, x_{\max}]$ are called the overload areas, and the quantisation error introduced is the overload error [108]. The granular error is the component that only accounts for quantisation noise within $[x_{\min}, x_{\max}]$ (see Figure 3.3).

Note that, the uniform quantisation is only optimal (in the minimum MSE sense) for a uniform distribution, where the MSE, σ_q^2 can be calculated by [108]

$$\sigma_q^2 = \sum_{q=1}^{N_i} \int_{b_{i-1}}^{b_i} (y_c - y_q)^2 f_X(x) dx. \tag{3.10}$$

From (3.10), it is known that the smaller the decision intervals, the smaller the error term $(y_c - y_q)^2$, thus σ_q^2 is inversely proportional to the number of decision intervals N_i. Given that this value is fixed, the quantiser design should find the placement of decision boundaries and

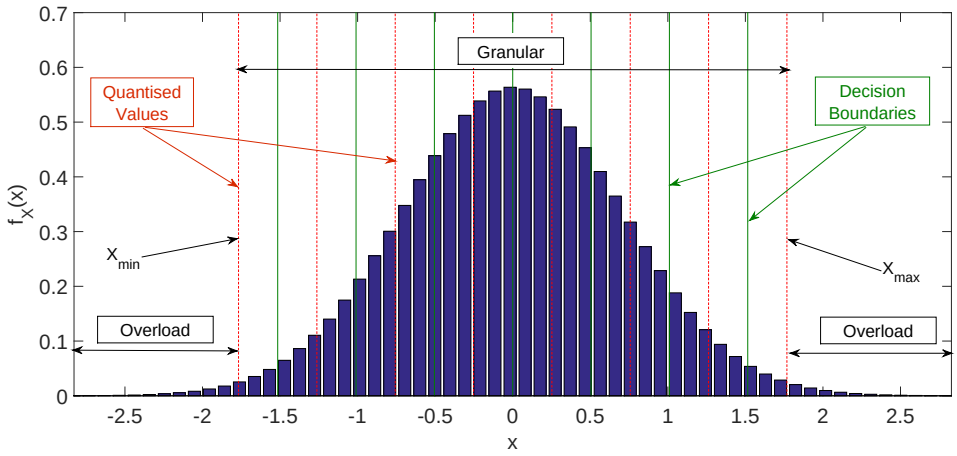


Figure 3.3: Histogram of $x = R_e\{y_c^2\}$.

quantised values so that σ_q^2 is minimised.

Although a vector quantisation performs better than scalar quantisation, this reduction comes at the expense of more computational complexity, hence, for potentially large virtual antenna array systems low complexity processing are desirable. In this chapter, it is considered that N_i quantisation levels, q_i, fall in the middle of the boundary points. Thus, if either the real or imaginary part of the signal received at each relay falls between the boundaries $b_i - 1$ and b_i, then it gets assigned a quantised value, q_i. Moreover, the quantisation range, $[l_b, u_b]$, is chosen as the one with the lowest MSE and depends on C. Due to the symmetric distribution of the noise-free constellation, it is assumed that $l_b = -u_b$. In general, when a uniform scalar quantiser is employed, it necessary to allocate $\log_2(N_i)$ bits to store each symbol.

3.4.1.3 Lloyd-Max Quantiser

In order to compensate for the distortion introduced by a suboptimal uniform scalar quantiser, multiple techniques can be employed; from increasing the number of bits per sample to implementing a non-uniform scalar quantiser such as Lloyd-Max or LBG [107], or even concatenating source coding, e.g. Huffman coding, to the quantisation process. In this subsection, the contribution of each error to the MSE weighted by its PDF is considered by employing the non-uniform Lloyd-Max scalar quantiser; which is an algorithm particularly attractive for real-time implementations due to its lower computational complexity [109]. The Lloyd-Max algorithm has previously been used for image processing [110], audio coding and signal processing [108]. In the context of digital communications and particularly for large MIMO antenna arrays, it can

be considered as an enabler for future wireless communications due to its low complexity and ability to take in to account the distribution of the received signals.

The Lloyd-Max quantiser (LM-SQ) satisfies the conditions that i) the quantised value for each decision interval is the centroid of the probability mass in the interval and ii) the decision boundary is simply the midpoint of the neighbouring quantised value [108]. Thus, given a PDF, $f_X(x)$ and a fixed number of decision intervals, N_i, it is required to find the set of boundaries $\{b_q\}_0^{N_i-1}$ and set of quantised values $\{y_q\}_0^{N_i-1}$ such that

$$y_q = \frac{\int_{b_q}^{b_{q-1}} x f_X(x) dx}{\int_{b_q}^{b_{q-1}} f_X(x) dx}, \tag{3.11}$$

and

$$b_q = \frac{1}{2}(y_q + y_{q+1}). \tag{3.12}$$

In order to solve (3.11) and (3.12) simultaneously, an iterative procedure is described in Algorithm 1 [108], where Δ is a positive variable accounting for the increment step and ϵ is an arbitrary threshold.

Algorithm 1 Lloyd-Max quantiser per dimension

1: Initialize $b_0 = -\infty$, $b_{N_i} = \infty$, $y_1 = y_1^{NF}$, $\theta = \epsilon$
2: **while** $|\theta| \geq \epsilon$ **do**
3: **for** $q = 1 : N_i$ **do**
4: Using y_q and b_{q-1} into (3.12) solve for b_q.
5: Using y_q and b_q into (3.11) solve for y_{q+1}.
6: **end for**
7: Calculate $\theta = y_{N_i} - \frac{\int_{b_q}^{b_{q-1}} x f_X(x) dx}{\int_{b_q}^{b_{q-1}} f_X(x) dx}$
8: **if** $\theta > 0$ **then**
9: $y_1 = y_1 - \Delta$
10: **else**
11: $y_1 = y_1 + \Delta$
12: **end if**
13: **end while**

From Algorithm 1, the LM-SQ could be seen as a special case of V-VQ, a one-dimensional vector quantiser which performs a local optimization per dimension by using the LBG algo-

rithm, where the real and imaginary components of y_c are independently quantised and achieve an optimal minimum MSE per dimension. For example, assuming that a 256-QAM constellation is transmitted, then the real (right) and imaginary (left) components of the received signals are shown in Figure 3.4. As can be seen, the distribution of the magnitude of the received signals is not uniform, but is normally distributed around zero. Thus, if a U-SQ is employed at the receiver the equally spaced partition boundaries will affect the MSE performance of the system. An a improved BLER performance would be observed when the LM-SQ algorithm is employed to define the decision boundaries and scalar quantised values accordingly to the PDF distribution, as the higher the probability of occurrence the smaller the step size is assigned to that region.

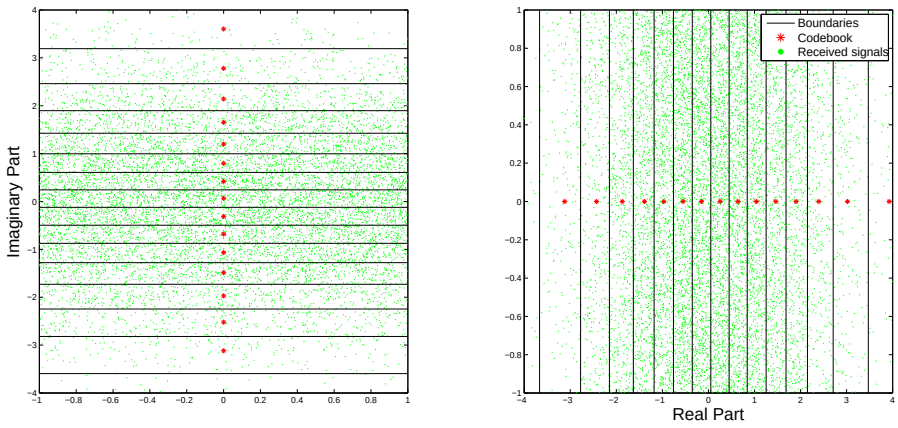
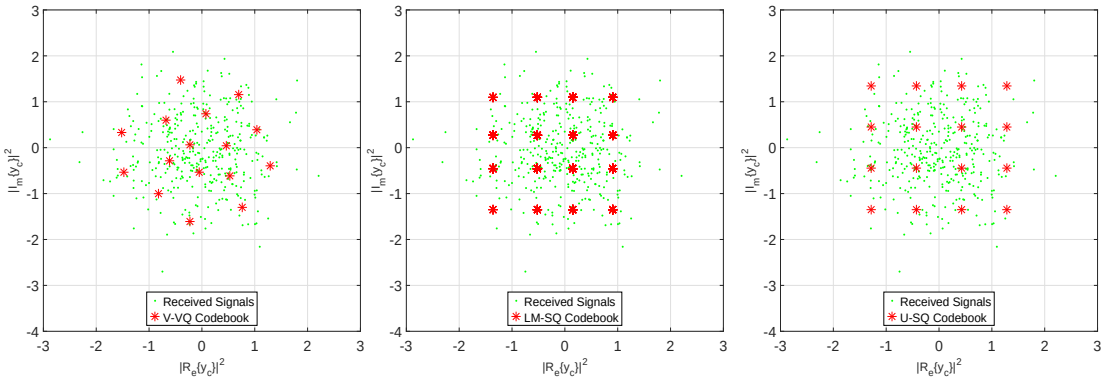


Figure 3.4: Decision boundaries for the real (right) and imaginary (left) components of a 256-QAM constellation, when the LM-SQ is employed at each relay.

Finally, it is assumed that the transmitter uses a 16-QAM constellation. An instantaneous version of y_c is shown in Figure 3.5. With the knowledge of the noise free constellation, either a LM-SQ or U-SQ approach can be used. The codebook designed by the V-VQ is also shown in Figure 3.5(a) for comparison. Even though the scalar quantiser implemented at the relay are suboptimal in terms of MSE, they are much simpler to design and execute and can achieve a comparable performance to the optimal V-VQ, as will be shown in Section 3.5 .

Figure 3.5 shows the codebooks designed by (a) V-VQ, (b) LM-SQ and (c) U-SQ schemes. As can be seen, given a training sequence, V-VQ generates a 2-dimensional optimum codebook. Meanwhile, LM-SQ optimizes the MSE per dimension by considering the distribution of the signal. Note that, given the symmetry of the M-QAM modulation and y_c^{NF} so is the

codebook obtained by using U-SQ, making this even simpler to design. Thus, even though the proposed schemes are suboptimal, they have a low computational complexity as shown in the next section.



(a) (b) (c)

Figure 3.5: 2-dimensional codebook design at c-th relay obtained by using the (a) V-VQ, (b) LM-SQ and (c) U-SQ quantisation schemes, with $C = 2$ bits per dimension/sample and 4-QAM modulation order, where the x- and y-axis represent the $|R_e\{y_c\}|^2$ and $|I_m\{y_c\}|^2$, respectively.

3.4.2 Complexity Analysis

Once the codebook has been designed by each relay node, it uses a look up table to assign a code to each received signal. In order to fairly compare the complexity of V-VQ, U-SQ and LM-SQ quantisers, in this section, the computation times for designing the codebook and encoding the received signal, and the overhead bits required for codebook transmission are presented.

3.4.2.1 Codebook design complexity

Recall that both V-VQ and LM-SQ use the LBG algorithm to design the codebook with a different size of training sequence. While V-VQ performs a two dimensional vector optimization, LM-SQ optimizes the MSE per dimension. Therefore, their design complexity mainly comes from the LBG algorithm implementation. In [111], the computational time for implementing the LBG algorithm is given by

$$T_{\text{LBG}} = I_s C_b V_s T_d + I_s (C_b - 1) V_s T_c, \tag{3.13}$$

where I_s is the number of iterations needed to meet the stopping criterion, C_b the codebook size and V_s the number of training vectors. The scalars T_d and T_c stand for the computational times for finding the distance between two vectors and comparing two distortion values, respectively. For V-VQ, which implements the LBG algorithm using as training sequence the noise-free constellation and considering that $I_s \leq V_s/C_b$ [111], then, the design time can be described as

$$T_{d,\mathrm{VQ}} \leq \frac{M^{2N_t}}{2^{2C}}(2^{2C}T_d + (2^{2C}-1)T_c). \tag{3.14}$$

On the other hand, for the proposed LM-SQ quantisation scheme, the computation time is given by

$$T_{d,\mathrm{LM}} \leq 2\frac{\sqrt{M}^{2N_t}}{2^{C}}(2^{C}T_d + (2^{C}-1)T_c). \tag{3.15}$$

A justification for (3.15) is as follows, in accordance with the LM-SQ algorithm in subsection 3.4.1.3, the real and imaginary parts of the received signal y_c are quantised independently at each relay. Thus, the number of codevectors required to reconstruct the M-QAM constellation is 2^C rather than 2^{2C}. During the initial stage, the LM-SQ algorithm is able to further reduce complexity by employing training scalar values instead of vectors. Without loss of generality and considering that the computational times for finding the distance between two vectors and comparing its distortion depend on the size of the vector, then T_d and T_c for V-VQ are greater that the ones required for LM-SQ.

On the other hand, when a U-SQ is considered the design time can be obtained by

$$T_{d,\mathrm{USQ}} \leq 2(\sqrt{M}N_tT_d), \tag{3.16}$$

where a square M-QAM constellation is assumed. Recall that, U-SQ is not only a non-iterative algorithm but also avoids using a training sequence. Thus, the design complexity is reduced to compute the maximum absolute value of the distribution points in each dimension. Even though a training sequence is not necessary, the availability of this information at the relays allows them to select the value u_b with lowest MSE. Without loss of generality, when a large V-MIMO system is considered, the complexity order of the design times are $T_{d,\mathrm{VQ}} = \mathcal{O}(M^{N_t})$, $T_{d,\mathrm{LM}} = \mathcal{O}(\sqrt{M}^{N_t})$ and $T_{d,\mathrm{USQ}} = \mathcal{O}(\sqrt{M}N_t)$, respectively. Comparing the complexity for V-VQ with the proposed schemes, the LM-SQ and U-SQ methods considerably reduce the

design complexity.

3.4.2.2 Encoding complexity and overhead

With a reliable codebook, each relay node quantises the received symbols by finding the codevector with lowest Euclidean distance to the received symbol. Particularly, the encoding algorithm for a V-VQ can be viewed as an exhaustive search algorithm. For a codebook of size 2^{2C}, the codevector selection for one symbol requires 2^{2C} distortion evaluations and $(2^{2C}-1)$ comparisons. Thus, the required time to search the codebook for one symbol can be written as [101]

$$T_{s,\mathrm{VQ}} = 2^{2C}T_d + (2^{2C}-1)T_c. \tag{3.17}$$

For the scalar quantisation schemes, the search requires 2^C evaluations and (2^C-1) comparisons per dimension. Therefore, the computation time for encoding each symbol is

$$T_{s,\mathrm{LM}} = T_{s,\mathrm{USQ}} = 2(2^C T_d + (2^C-1)T_c). \tag{3.18}$$

The computation time in (3.18) represents the worst case for the proposed schemes. Nonetheless, a much simpler search can also be performed by employing a slicing operation per dimension. Finally, the overhead incurred by the quantisation techniques is considered by calculating the number of bits required to transmit the codebook to the destination node, task performed every L-consecutive symbol periods. For the optimal V-VQ the whole codebook needs to be transmitted. Considering the dimension of each codevector, the codebook size and assuming that B-bits are used to represent each codevector, the overhead bits are

$$O_{h,\mathrm{VQ}} = 2^{2C+1}B/L. \tag{3.19}$$

Meanwhile, the proposed schemes only require the information per dimension to reconstruct the codebook, thus,

$$O_{h,\mathrm{LM}} = 2^{C+1}B/L, \tag{3.20}$$

for LM-SQ and

$$O_{h,\mathrm{USQ}} = 3B/L, \tag{3.21}$$

in the case of U-SQ, which only requires the knowledge of the value u_b per dimension and C, reducing significantly the total overhead.

3.5 Numerical results

In this section, the block error rate (BLER) performance and complexity analysis of the V-MIMO system considering low complexity quantisation techniques at the relay nodes are presented. Unless stated otherwise, the results shown in this section pertain to a BICM-MIMO system with 4-QAM Gray mapping modulation and a Rate-1/2 convolutional code with constraint length 7 and generator polynomials $[133, 171]_{\text{octal}}$. Rayleigh block - fading with 10^5 blocks is assumed, where each block has 200 consecutive symbol periods. Moreover, the LLRs are computed through the max-log approximation [103] and ZF detection is performed at the destination node. The V-VQ is a two-dimensional vector quantiser with bit rate of 2 bits per dimension/sample; further, the total transmitted power E_s is normalized to unity and $N_c = N_r - 1$ is the number of relays employed. The simulation results are computed through the Monte Carlo method.

3.5.1 BLER evaluation of the Codebook design

Recall, that for a fixed number of decision intervals, the smaller the quantisation step is, the smaller the granular quantisation noise becomes. However, the smaller quantisation step size translates into a smaller u_b, which leads to larger overload areas, hence a larger overload quantisation error [108]. For our numerical simulations, u_b was selected according to Figure 3.6, which shows the MSE obtained for $C = 2, 4$ and 8 bits per dimension/sample.

As mentioned above, the codebook design complexity and encoding complexity of the V-VQ quantiser is very high. To reduce the complexity and enhance the practicality of the CF cooperation, it is proposed to employ the Lloyd-Max algorithm per dimension to design the codebook at the relay. Figure 3.7 compares the BLER performance for 10×20 ($N_c = 19$) and 10×30 ($N_c = 29$) V-MIMO systems when different quantisation techniques are employed. It can be seen that, when $N_r = 20$ the performance gap is about 0.38 dB (LM-SQ) and 1.67 dB (U-SQ) from the optimal V-VQ at BLER $= 10^{-2}$. For the 10×30 case, the losses are 0.30 dB and 0.83 dB, respectively. From Figure 3.7, it can be noticed that increasing the number of receiver antennas not only improves the system's performance but also reduces the gap between V-VQ

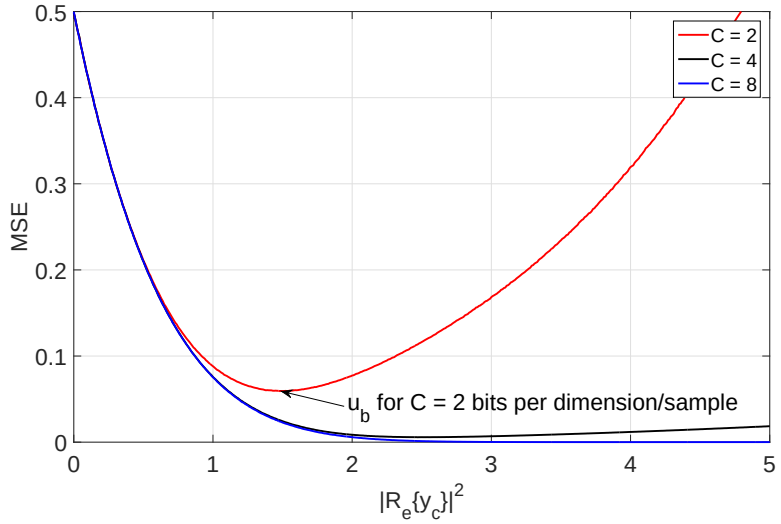


Figure 3.6: Minimum square error (MSE) obtained wtih $C = 2, 4$ and 8 bits per dimension/sample when a 4-QAM modulation order is considered.

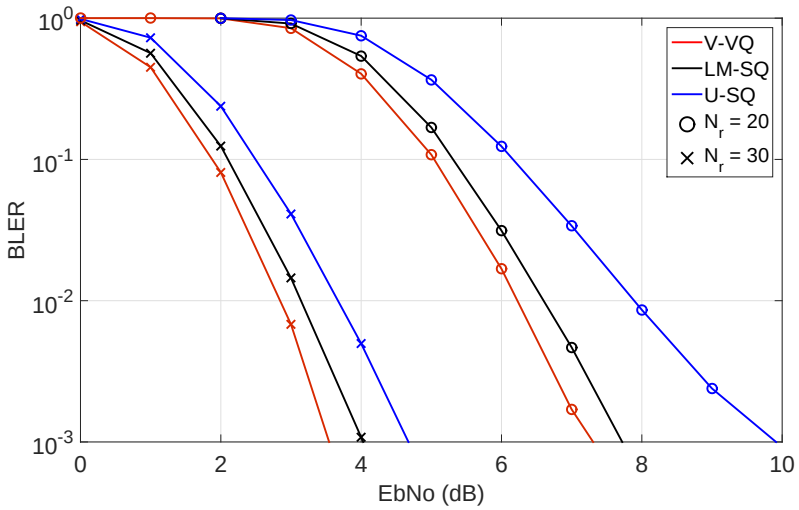


Figure 3.7: BLER performance of V-MIMO system with $N_t = 10$, $N_r = 20$ and 30, 4-QAM modulation and $C = 2$ bits per dimension/sample.

and the scalar quantisation techniques.

It is of particular interest to analyse the performance of large V-MIMO systems, thus, Figure 3.8 compares the BLER performance when $N_t = 50$, and $N_r = 100$ and 150 are considered. For the 50×100 system, LM-SQ and U-SQ perform within 0.54 dB and 1.65 dB of V-VQ, respectively. As the number of receivers is increased to $N_r = 150$, the proposed quantisation techniques perform better, requiring 0.29 dB and 0.79 dB higher E_b/N_0 than V-VQ, respectively. By comparing Figures 3.7 and 3.8, it can be seen that the performance mainly depends

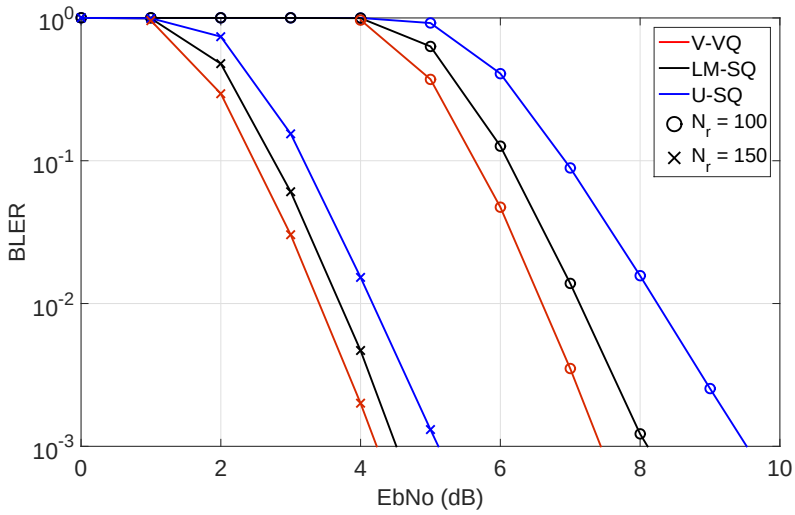


Figure 3.8: BLER performance of V-MIMO system with $N_t = 50$, $N_r = 100$ and 150, 4-QAM modulation and $C = 2$ bits per dimension/sample.

on the ratio N_r/N_t.

Continuing with the BLER performance analysis, Table 3.1 shows the performance gain of LM-SQ over U-SQ for different system configurations at BLER $= 10^{-2}$. The number of receivers was selected to get the same ratios, N_r/N_t, when $N_t = 10, 20, 50$ and 100 are employed. From Table 3.1, it can be seen that (a) as N_r increases the performance gap between LM-SQ and U-SQ reduces and, (b) the performance gap between the proposed schemes is similar for systems with different configurations but the same ratio N_r/N_t.

N_t	N_r	LM-SQ	U-SQ	E_b/N_0 gap (dB)	N_r/N_t
10	15	11.67	18.90	7.23	1.5
	20	6.60	7.89	1.29	2.0
	30	3.14	3.67	0.53	3.0
20	30	11.46	18.75	7.29	1.5
	40	6.72	7.93	1.21	2.0
	60	3.29	3.85	0.56	3.0
50	75	12.05	20.09	8.04	1.5
	100	7.13	8.24	1.11	2.0
	150	3.70	4.20	0.50	3.0
100	150	13.43	> 20	≈ 9	1.5
	200	7.42	8.56	1.14	2.0
	300	3.91	4.48	0.57	3.0

Table 3.1: E_b/N_0 performance gap between LM-SQ and U-SQ quantiser when $C = 2$ bits per dimension/sample at BLER = 10^{-2} and 4-QAM modulation order system are considered.

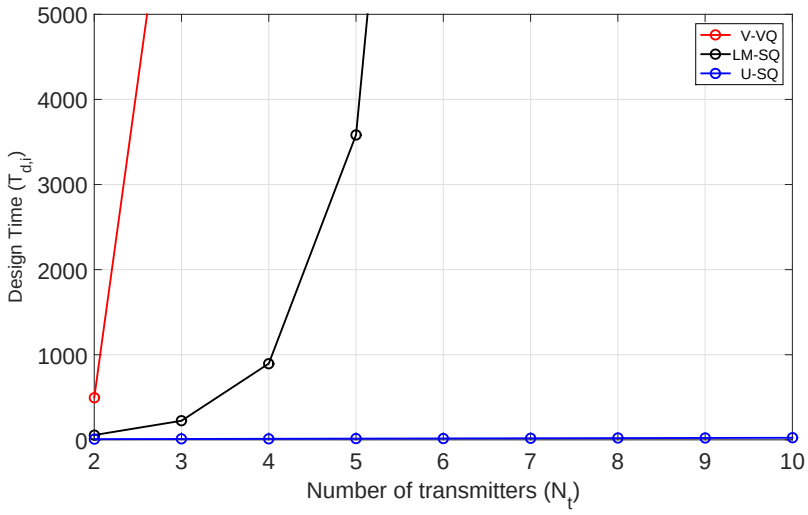


Figure 3.9: Computation times for designing a reliable codebook with $C = 2$ bits per dimension/sample and 4-QAM modulation order.

3.5.2 Complexity evaluation of the Codebook design

Figures 3.9 and 3.10 compare the design and search computation times according to equations (3.15) and (3.16) respectively. Figure 3.9 shows the computation times required for constructing a reliable codebook when $C = 2$, $M = 4$ and where T_c and T_d are normalized to one for simplicity. As can be seen, as N_t increases, the design complexity for V-VQ grows exponentially according to $T_{d,\mathrm{VQ}} \leq 4^{2N_t-2}(16T_d + 15T_c)$. Even though, the proposed LM-SQ also shows an exponential behaviour, it has lower exponential growth rate as $T_{d,\mathrm{LM}} \leq 2^{2N_t-2}(4T_d + 3T_c)$. Meanwhile, the design complexity of U-SQ only grows linearly with N_t. Further, as shown in Figure 3.10, the computation time required for quantising the received symbols of LM-SQ and U-SQ ($8T_d + 6T_c$) is lower than that of V-VQ ($16T_d + 15T_c$).

On the other hand, the overhead bits for transmitting the codebook for a given quantisation rate of 2 bits per dimension/sample is computed considering $L = 200$ consecutive symbol periods and $B = 8$ bits per element. The V-VQ requires 1.28 bits/period which is four times greater than 0.32 bits/period required by LM-SQ and much larger than that of U-SQ requiring 0.12 bits/period. Since, the proposed quantisation schemes are more efficient in complexity and achieve a comparable performance to the optimal V-VQ, they could be considered as a suitable alternative to enable a large-scale cooperative wireless network.

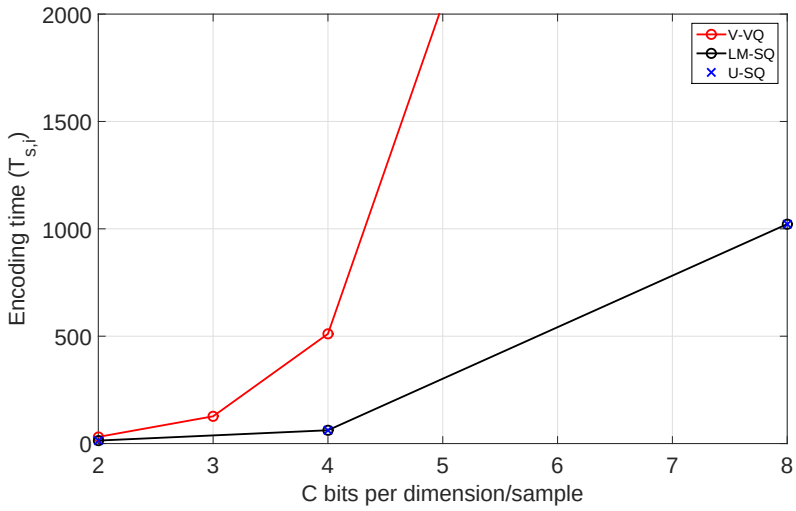


Figure 3.10: Computation times for quantising a symbol with $C = 2$ bits per dimension/sample and 4-QAM modulation order.

3.6 Summary

In this chapter, a cooperative virtual MIMO system that uses a multi-antenna transmitter and implements BICM transmission and CF cooperation among closely spaced single-antenna receivers was presented. Since constructing a reliable codebook is the most computationally complex task performed by the relay nodes, two codebook design alternatives to the optimal Voronoi VQ were introduced. A comparison in terms of the codebook design complexity, encoding complexity and the overhead bits required for transmitting the codebook to the destination node was also presented.

Furthermore, from practical considerations the performance of large V-MIMO systems was analysed. An evaluation of the BLER performance when $N_t = 50$ and $N_r = 100$ and 150 was presented. For the 50×100 V-MIMO system, simulation results have shown that, the LM-SQ and U-SQ perform within 0.54 dB and 1.65 dB of V-VQ, respectively. To illustrate the benefit of CF cooperation in very large virtual-MIMO systems, the number of receivers was increased up to $N_r = 150$. It was shown that the proposed quantisation techniques required 0.29 dB and 0.79 dB higher E_b/N_0 than V-VQ, respectively. Moreover, simulation results showed that the BLER performance mainly depends on the ratio of the number of antennas at the transmitter and the number of single-antenna devices at the receiver. In addition, for a given quantisation rate of 2 bits per dimension/sample and $B = 8$ bits per element. The V-VQ requires 1.28 bits/period which is four times greater than 0.32 bits/period required by LM-SQ and much larger than that of U-SQ requiring 0.12 bits/period. Since, the proposed quantisation

schemes are more efficient in terms of computational complexity and achieve a comparable performance to the optimal V-VQ, they can be considered as a suitable alternative to enable large-scale cooperative wireless networks.

Chapter 4
Energy Efficient Massive MIMO Wireless Networks

In this chapter, a novel cooperative virtual multi-user MIMO (MU-MIMO) system that uses a large multi-antenna transmitter and implements bit-interleaved coded modulation (BICM) transmission and decode-and-forward (DF) relay cooperation among multiple receiving nodes is presented. In DF-based cooperation, decoding, re-encoding and forwarding the estimated received signals to the destination node for subsequent processing are the main tasks that each relay has to perform. Even though large multi-antenna approaches have been identified as an effective way to enhance the spectral efficiency (SE) of wireless networks, this comes at the expense of linearly increasing the power consumption due to the use of multiple radio frequency (RF) chains to support the antennas. Thus, in this chapter, the energy efficiency and the effective throughput of the proposed DF cooperation system are analysed to understand its advantages and trade-offs.

In order to quantify any energy efficiency (EE) gains achieved by the cooperative wireless system a common framework is required. Thus, a power consumption model that considers not only the power consumed by the transmitter but also that takes into account the overhead power consumption incurred by the relays is presented. Moreover, the impact of the imperfect CSI on the throughput and energy efficiency of the system is studied. Finally, a power allocation algorithm is used to reduce the total power consumption.

Simulation results show that, in practical scenarios where a number of users share the radio spectrum and thus there is a restriction on the available resource blocks, cooperative systems can be employed to meet the demand for high data rates. Moreover, it is shown that wireless cooperation between users is more energy efficient than using high modulation order transmission in the low resource block usage case. Numerical results show that the larger the number of transmit antennas, the lower the impact of imperfect CSI on the system's performance.

The remainder of this chapter is organised as follows. Section 4.1 presents the background and motivation of this chapter. The system model is described in Section 4.2. The power

consumption model and the analysis of the spectral and energy efficiency of the virtual MU-MIMO are presented in Section 4.3. Simulation results are shown in Section 4.4. Finally, conclusions are presented in Section 4.5.

4.1 Introduction

Due to the forthcoming spectrum shortage and the increasing demand for data services, the design of current wireless networks has been focused on enabling high spectral efficiency systems [112]. Virtual MIMO (V-MIMO) and Massive MIMO approaches have been put forward as effective means to enhance the SE performance [61, 102]. However, this comes at the expense of linearly increasing circuit power consumption due to the use of multiple RF chains to connect the antennas [113]. Since the requirement for ubiquitous access has significantly increased the energy demands [12,112], the study and development of techniques that can provide highly efficient energy wireless systems have recently attracted significant attention.

Previous works have been focused on studying the EE of either V-MIMO or Massive MIMO systems, but not both. As an example, the authors in [14, 114, 115] have studied the energy efficiency of a standard relay channel, i.e., a three-node network where the source transmits to the destination with the assistance of the relay. Meanwhile, the EE performance of a relay channel corrupted by additive white Gaussian is bounded in [114]. The authors in [115] investigated the EE performance of a relay channel operating in the time division duplex (TDD) when a low-power regime is considered. Additionally, the normalized achievable minimum energy-per-bit that enables optimal power allocation between the source and the relay was derived in [14]. Nevertheless, in these papers the circuit energy consumption is not considered, which may become significant in practical wireless systems when the transmit antenna array is very large.

Assuming that a space-time block coding (STBC) scheme is used by the transmitter, the energy and delay efficiencies of a MIMO cooperative wireless sensor network were derived in [116] using semi-analytic techniques. The results showed that with judicious choice of design parameters the virtual MIMO technique can provide significant improvement in terms of energy efficiency and transmission delay. Further, in [55] an upper bound of EE as a function of SE for a V-MIMO system with CF cooperation is presented and based on this an EE optimization problem is formulated. The achievable rate for the Gaussian MIMO relay channel that

implements zero-forcing (ZF) precoding and the DF protocol was obtained in [117], by using standard convex optimization techniques.

Furthermore, from the Massive MIMO perspective, several studies have been conducted to analyse the way that a large number of transmit and receive antennas impact the EE performance. In [113], the performance of a point-to-point large-scale MIMO systems that considers the overall power consumption on both the transmitter and the receiver was presented and the optimal RF chain configuration to maximize the transmission rate under power constraint was derived. Meanwhile, in [118], the authors determined the minimum required transmit power for a given outage probability, showing that antenna selection is in general the most energy efficient option as it requires a single radio-frequency chain. Additionally, the authors in [119] presented the EE for the uplink MU-MIMO configuration and an optimal power allocation strategy while both RF transmission power and device electronic circuit power were considered. An improved circuit management scheme showed that user antennas should be used only when the corresponding spatial channels are sufficiently good and using them improves the overall EE of the network.

Generally speaking, there are mainly two ways to reduce energy consumption: 1) by using energy aware components in the base stations and 2) by implementing energy aware network deployment strategies. In order to quantify the gains achieved by employing energy aware techniques, an appropriate energy efficiency metric must be defined [120]. In [121], a power consumption model for a LTE user equipment was presented. The model included functions of UL and DL power and data rate where realistic measurements from handset devices for power consumption were used. Moreover, a power consumption model for mobile broadband access networks that takes into account the backhaul, and the main trade-offs between infrastructure, energy, and spectrum costs when low-power micro base stations are employed was introduced in [112]. In [120], the authors developed power models for macro and micro base stations relying on data sheets of several Global System for Mobile Communications (GSM) and Universal Mobile Telecommunications Service (UMTS) base stations with a focus on the component level, e.g., power amplifier and cooling equipment.

Since the effect of circuit power consumption is more serious when the transmitter is equipped with a massive number of antennas [112], different models for power consumption that consider the fundamental power for operating the circuit at the transmitter have been proposed. In [122], a power consumption model that considers not only the transmit power on the power amplifier

but also the circuit power dissipated by analogue devices and other loss factors in base stations was presented. Simulation results showed that the energy efficiency becomes a quasi-concave function as the number of antennas increases. Moreover, a realistic power consumption model that reveals how the number of antennas, active users, and transmit power jointly affect the EE was presented in [7, 123]. It was shown that the transmit power increases with the number of antennas. This implied that energy-efficient systems can operate in high signal-to-noise ratio regimes in which interference-suppressing signal processing is mandatory.

Furthermore, the authors in [9] presented a power model that supports a broad range of network scenarios and base station types, features and configurations, including base station deactivation and scaling factors which are used to predict the power consumption of base stations up to the year 2020.

In this chapter, the energy efficiency and throughput for a V-MIMO system that uses a large transmit antenna array and implements DF cooperation between the user terminals are studied. The main contributions of this chapter can be summarised as follows: i) the total power consumed considering not only the transmit and circuit power consumption but also the overhead power consumed by the relays is computed, ii) the energy efficiency and the effective throughput of a cooperative system, when the available resource blocks are limited, are evaluated, iii) the impact of imperfect CSI on the system's performance is evaluated and a power allocation algorithm to reduce power consumption is implemented. Simulation results show that the optimal throughput is obtained when there is no restriction on the use of resource blocks. However, in a practical scenario when a number of users are sharing the frequency resources, cooperative systems could be implemented in order to meet the demand for data rates. Such wireless cooperation between users is more energy efficient than using a high modulation order transmission in the low resource block usage case. Moreover, it is shown that the larger the number of transmit antennas, the lower the impact of imperfect CSI on the system's performance.

4.2 System Model

Consider the downlink transmission of a cooperative virtual MU-MIMO OFDM system with a N_t-antenna base station (BS) and N_r ($\ll N_t$) single-antenna user terminals (UEs), as shown in Figure 4.1. The multicarrier system, where a block of N_{rb} information symbols is transmitted in parallel on N_{rb} subcarriers, can be efficiently implemented in discrete time using an inverse

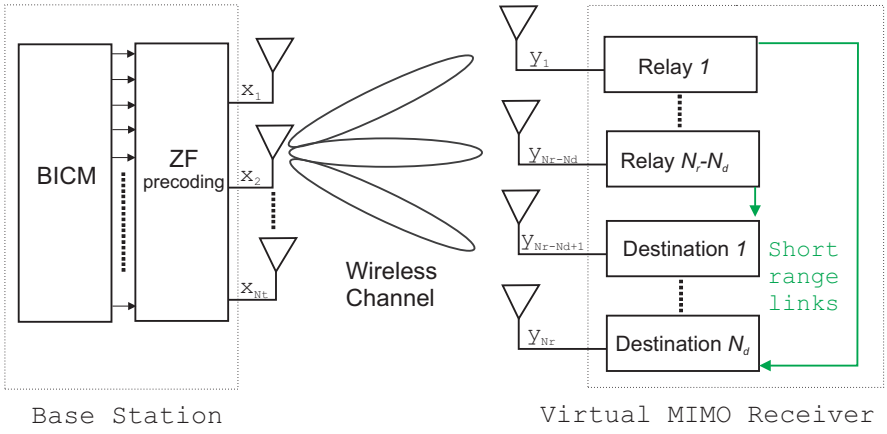


Figure 4.1: Block diagram of MIMO base station and virtual MU-MIMO receiver.

fast Fourier transform (IFFT) to act as a modulator and an FFT to act as a demodulator [124]. Similar to Chapter 3, at the transmitter side, a BICM scheme is employed to provide error correction and improve the system's performance. The data bits are encoded using a rate-R_b linear convolutional encoder and then passed through an ideal bitwise interleaver, which rearranges the coded bits using a random permutation. Next, the interleaved coded bits are demultiplexed into N_t streams. In each stream $k = 1, \ldots, N_t$ a group of m_s^b coded bits are mapped to a data symbol s_k from a symbol alphabet $\mathcal{S}$ of size $|\mathcal{S}| = 2^{m_s^b}$. Thus, the transmission rate is $N_t m_s^b R_b$ bits per channel use. It is further assumed that the BS uses linear precoding techniques to process the signal before transmitting to all users. In order to guarantee a practical implementation and according with Section 2.1.4, it is considered that the BS and UEs operate according to the TDD protocol. Hence, the time-frequency resources required for CSI are independent of the number of transmit antennas and the uplink pilots enable the BS to estimate the UE channels. In this work, perfect hardware chain calibration is considered and thus the uplink and downlink channels are considered reciprocal. Moreover, a block fading model, where the small-scale fading vector remains constant during a coherence block of L symbols and is independent in different coherence blocks, is assumed. Furthermore, it is assumed that the UEs are closely located with non-light-sight (NLoS) conditions, where the rich scattering allows spatial separation of the intended signal for each UE [125]. Figure 4.1 shows the diagram of a cooperative virtual MU-MIMO system that uses DF cooperation among multiple receiving nodes to increase the capacity of the destination ones. It is further assumed that the relays and destination nodes cooperate through one-shot conference links and that each destination node is equipped with a traditional receiver which observes the signals from the transmitter, and an intra-cluster receiver which observes signals from the relays. Section 4.2.3 provides the reader

with further information regarding the wireless cooperation carried out at the receiver side.

4.2.1 Uplink Training

Let τ_u be the number of symbols per coherence interval used for uplink pilots and transmitted simultaneously at each UE. Since the pilot sequences of the N_r UEs are pair-wise orthogonal, then, it is required that $\tau_u \geq N_r$. Considering that the BS serves the UEs over a slow flat-fading channel and the channel estimation error can be treated as a component that contributes as an additional source of distortion independent of the noise component, then the minimum mean-square error (MMSE) estimate of the channel matrix, $\mathbf{H} \in \mathbb{C}^{N_t \times N_r}$ is given by [126]

$$\hat{\mathbf{H}} = \frac{t_u P_u}{t_u P_u + 1}\mathbf{H} + \frac{\sqrt{t_u P_u}}{t_u P_u + 1}\mathbf{N}_u, \tag{4.1}$$

where P_u denotes the average transmit power of each uplink pilot symbol and $\mathbf{N}_u \in \mathbb{C}^{N_t \times N_r}$ represents the distortion introduced by the imperfect channel estimation with entries $n_u \sim \mathcal{CN}(0, 1)$. Then the channel matrix is equivalent to

$$\mathbf{H} = \hat{\mathbf{H}} + \mathcal{E}, \tag{4.2}$$

where $\mathcal{E}$ is the channel estimation error. Since MMSE channel estimation is assumed, $\hat{\mathbf{H}}$ and $\mathcal{E}$ are independent [127] and their elements are given by $\hat{h}_{i,j} \sim \mathcal{CN}\left(0, \frac{t_u P_u}{t_u P_u + 1}\right)$ and $\varepsilon_{i,j} \sim \mathcal{CN}\left(0, \frac{\sqrt{t_u P_u}}{t_u P_u + 1}\right)$, respectively.

4.2.2 Downlink Transmission

The BS simultaneously transmits the modulated data symbols through N_t antennas using a transmit vector, $\mathbf{x} = [x_1, \ldots, x_{N_t}]^{\mathrm{T}}$, that satisfies the power constraint $\mathbb{E}[||\mathbf{x}||^2] = P$. The BS uses the channel estimate $\hat{\mathbf{H}}$ to linearly precode the symbols, i.e., $\mathbf{x}$ is a linear combination of the symbols intended for the N_r UEs. Let s_k be the symbol to be transmitted to the k-th user, with $\mathbb{E}[|s_k|^2] = 1$. Then, the linearly precoded vector $\mathbf{x}$ can be written as

$$\mathbf{x} = \sqrt{p_d}\mathbf{F}\mathbf{s}, \tag{4.3}$$

where $\mathbf{s} \in \mathbb{C}^{N_r \times 1}$ is the vector of data symbols, $\mathbf{F} \in \mathbb{C}^{N_t \times N_r}$ is the precoding matrix and p_d is a normalization constant chosen to satisfy the power constraint $\mathbb{E}[||\mathbf{x}||^2] = P$. Thus,

$$p_d = \frac{P}{\mathbb{E}\{\text{tr}(\mathbf{F}\mathbf{F}^H)\}}. \tag{4.4}$$

Collectively, the received signal vector of the N_r UEs can be written as

$$\mathbf{y} = \mathbf{H}^T\mathbf{x} + \mathbf{n} = \sqrt{p_d}\mathbf{H}^T\mathbf{F}\mathbf{s} + \mathbf{n}, \tag{4.5}$$

where $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the additive noise vector for the UEs with $n_k \sim \mathcal{CN}(0, \sigma_N^2)$. Aiming to null out the multiuser interference by projecting each stream onto the orthogonal complement of the inter-user interference, zero-forcing (ZF) precoding is adopted. Thus, the precoding matrix $\mathbf{F}$ can be expressed as [126]

$$\mathbf{F} = \hat{\mathbf{H}}^*(\hat{\mathbf{H}}^T\hat{\mathbf{H}}^*)^{-1}, \tag{4.6}$$

where $(\cdot)^*$ and $(\cdot)^T$ denote the matrix conjugate and transpose operations, respectively. Finally, the received signal at the k-th UE is given by

$$\begin{aligned} y_k &= \sqrt{p_d}\hat{\mathbf{h}}_k^T\mathbf{F}\mathbf{s} + n_k \\ &= \sqrt{p_d}\hat{\mathbf{h}}_k^T\mathbf{f}_k s_k + \sum_{j \neq k} \sqrt{p_d}\hat{\mathbf{h}}_k^T\mathbf{f}_j s_j + n_k, \end{aligned} \tag{4.7}$$

where $\mathbf{h}_k$ and $\mathbf{f}_k$ are the k-th column of the channel and precoding matrix, respectively. Even though linear precoder schemes are not optimal, they can be used to reduce signal processing complexity. Moreover, when the number of antennas used at the transmitter is large enough, then the performance of linear precoders is nearly-optimal [61, 128].

4.2.3 Cooperative Virtual MU-MIMO System

At the receiver side, a cooperative virtual MU-MIMO system that uses DF cooperation among multiple receiving nodes to increase the capacity of the destination nodes is considered. In other words, each of the N_d destination nodes is assisted by N_c relays to increase their capacity. Since

the cluster of $N_r = N_d(N_c + 1)$ receivers are assumed to be closely spaced, it is reasonable to expect that the wireless cooperation between the relays and destination is considerably better and more stable than that between the transmitter and receivers. In fact, it is highly likely that a high channel capacity could be achieved due to the high reliability of these short range links. The one-shot cooperation link requires the relay to decode the signals and send over to their destination node (see Figure 4.1).

In this scenario, it is assumed that each of the single-antenna receivers acts either as a relay or as a destination node. With a DF cooperative protocol, each relay decodes the received signal, re-encodes and retransmits the signal

$$\tilde{s}_k = \hat{s}_k, \tag{4.8}$$

where $\hat{s}_k$ is the relay's estimate of s_k. Assuming that the signal is decoded correctly, i.e., $\hat{s}_k = s_k$, then the signal received at the destination node can be written as

$$\tilde{y}_k = \sqrt{P_r} h_{rd} s_k + n_d, \tag{4.9}$$

where P_r denotes the maximum transmit power at each relay. The short-range cooperation link between the relay and the destination is modelled as an additive white Gaussian noise channel n_d. The power gain of the cooperation channel to the power gain of the data channel is represented by $g_{rs} = \text{SNR}_{\text{coop}}/\text{SNR}_{\text{data}}$, and it remains constant throughout the system (see Figure 4.2). Since a scenario where the UEs are closely located is assumed, then we are interested in high values of g_{rs} [55].

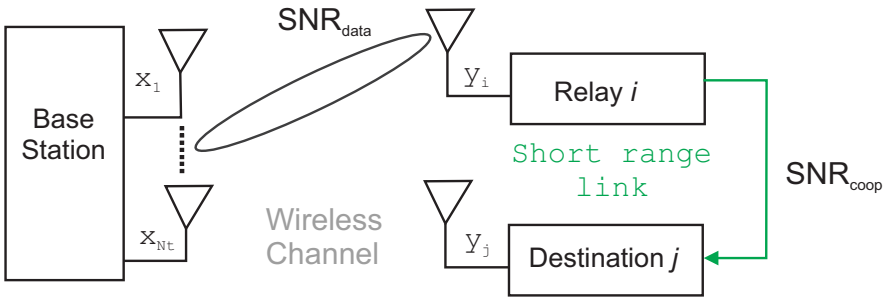


Figure 4.2: Simplified block diagram of the cooperative system, where SNR_{data} and SNR_{coop} represent the power gain of the data and cooperative channels, respectively.

Note that in this work, it was assumed that each destination node is assisted by the same number of relays. Nevertheless, it is also possible that this number varies according to the demand

in data rate at each destination node and the traffic load of the system. As mentioned in Section 3.2, this scenario is feasible if the orthogonality between the cooperative one-shot links is guaranteed, otherwise, the interference and complexity at the decoder would make the system impractical for implementation.

4.3 Energy Efficiency

The aim of this section is to quantify the overall power consumption of the virtual MU-MIMO system described in Section 4.2 and to analyse its energy efficiency performance. Thus, it is considered that the power expenditure in both the downlink and the cooperative links is the sum of the power consumed at the BS, the destination and relays nodes and a fixed constant accounting for the load-independent power consumption.

4.3.1 Power Consumption Model

In order to compute the power consumed by the virtual MU-MIMO system rather than only the transmitted power, the three main components from the power modelling point of view within the network are considered: (a) the multi-antenna BS, (b) the relay cooperation and (c) the load-independent power consumption.

For each base station within the network, the power consumption model should consider the five main power-drawing elements: the power amplifier (PA), the analogue front-end, the digital baseband, the control and network backhaul, and power systems [9]. Based on the models proposed in [7, 55, 129], the power consumption at the BS of a virtual MU-MIMO system is modelled as follows

$$P_{\text{BS}} = \frac{1}{\delta} P + N_t P_{\text{ct}} + m_s^b N_r P_{\text{cod}} + P_{\text{lo}}, \tag{4.10}$$

where δ denotes the efficiency of the PA, P is the total transmit power consumption and $m_s^b = \log_2(M)$ is the number of bits transmitted per data symbol, since it is assumed that s_k belongs to an M-QAM constellation. As described in [7], the power required to run the circuit components (including filters, converters and mixers) is denoted as P_{ct}, and it is linearly dependent on N_t. Further, P_{cod} is the power coding and scales with the number of user terminals N_r and

the modulation efficiency. Meanwhile, P_{lo} represents the power consumed by the single local oscillator.

The cluster of receivers, that includes the destination and relay nodes are assumed to be closely spaced and cooperate through one-shot links. Each relay is required to decode, encode and re-transmit the signal to its destination node. Thus, the power consumption at the receiver side should account for the circuit power consumption at the UEs, signal processing power consumption and the transmit power required to carry out the wireless cooperation. The total power consumed at the receiver side can be computed as

$$P_{\text{UE}} = N_r P_{\text{cr}} + N_c N_d P_r + m_s^b N_r N_d P_{\text{cod}} + m_s^b (N_r + N_c N_d) P_{\text{dec}}, \tag{4.11}$$

where P_{cr} is the circuit power consumption per UE and P_r accounts for the transmit power at each relay. Following the assumption in [55], the power allocation ratio between the relay and the transmitter is defined as $\gamma = P_r/P$. Thus, with the knowledge of P, the relay decides its own transmit power. The power for decoding and re-coding prior to the retransmission are, respectively, P_{dec} and P_{cod}, and depend on the number of bits processed. Moreover, a constant value P_0 accounting for the fixed power consumption required for load-independent power of the backhaul infrastructure, control signalling and site-cooling is considered [7]. Finally, the power consumption of the cooperative system can be written as

$$P_{\text{cons}} = P_{\text{BS}} + P_{\text{UE}} + P_0. \tag{4.12}$$

4.3.2 Achievable Rate

In this subsection, the achievable rate of the system introduced in Section 4.2 is described. Assuming that the channel can be estimated at the transmitter, the ergodic achievable rate at the k-th node is given by

$$R_k = \log_2(1 + \tau), \tag{4.13}$$

where τ is the associated signal-to-interference-plus-noise ratio (SINR) and can be calculated using equation (4.7). Considering equation (2.22) and after substituting the value of SINR into equation (4.13), the achievable rate at the k-th node can be written as

$$R_k = \begin{cases} \mathbb{E}\left[\log_2\left(1+\frac{P|\hat{\mathbf{h}}_k^T\mathbf{f}_k|^2}{\sum_{j\neq k}P|\hat{\mathbf{h}}_k^T\mathbf{f}_j|^2+\sigma_N^2}\right)\right], & \text{if direct link,} \\ \frac{1}{2}\min\left\{\mathbb{E}\left[\log_2\left(1+\frac{P|\hat{\mathbf{h}}_k^T\mathbf{f}_k|^2}{\sum_{j\neq k}P|\hat{\mathbf{h}}_k^T\mathbf{f}_j|^2+\sigma_N^2}\right)\right], \mathbb{E}\left[\log_2\left(1+P_r|h_{rd}|^2\right)\right]\right\}, & \text{otherwise} \end{cases} \tag{4.14}$$

where h_{rd} is the channel gain between relay and destination node and P and P_r are the transmit power at the transmitter and relay node, respectively. Notice that if the destination node is assisted by a relay, as in the lower part of (4.14), then the achievable rate is limited by the minimum rate between the base station-relay and relay-destination links according to equation (2.22). Finally, the sum achievable rate at the i-the destination node that is assisted by N_c relays can be computed as follows

$$R_i^{N_c} = \sum_{k=1}^{N_c}\left\{\frac{1}{2}\min\left\{\mathbb{E}\left[\log_2\left(1+\frac{P|\hat{\mathbf{h}}_k^T\mathbf{f}_k|^2}{\sum_{j\neq k}P|\hat{\mathbf{h}}_k^T\mathbf{f}_j|^2+\sigma_N^2}\right)\right], \mathbb{E}\left[\log_2\left(1+P_r|h_{rd}|^2\right)\right]\right\}\right\} + \mathbb{E}\left[\log_2\left(1+\frac{P|\hat{\mathbf{h}}_i^T\mathbf{f}_i|^2}{\sum_{j\neq i}P|\hat{\mathbf{h}}_i^T\mathbf{f}_j|^2+\sigma_N^2}\right)\right]. \tag{4.15}$$

4.3.3 Energy Efficiency Analysis

In this section, the energy efficiency of the system described in Section 4.2 is studied for three different scenarios. First, it is considered that a small number of UEs is served by the BS. Therefore, the BS is allowed to use more than one resource block per destination node to fulfil the target rate (See Figure 4.3). Since each resource block consists of one slot in both the time and frequency domains, when there is no restriction on the number of resource blocks, the k-th UE is served through N_{rb} carriers. Nevertheless, in a practical scenario, where the number of resource blocks per user terminal is restricted, different approaches can be considered to satisfy the UEs' demand for data rate. On one side, a cooperative wireless system, where the throughput is increased by the number of relays serving each destination node, can be implemented. On the other side, it is also possible to increase the modulation order per beam. In these two last scenarios, precoding allows multiple UEs sharing the same time-frequency resources. Figure 4.4 shows a graphic representation of two possible scenarios of the traffic load of the network. On the left, the network operates under low traffic load, so the available

resource blocks for the k-th user are not limited. On the right, the traffic load of the network has increased and thus, the BS is required either to select different time-resource blocks for transmission or to increase the modulation order per beam.

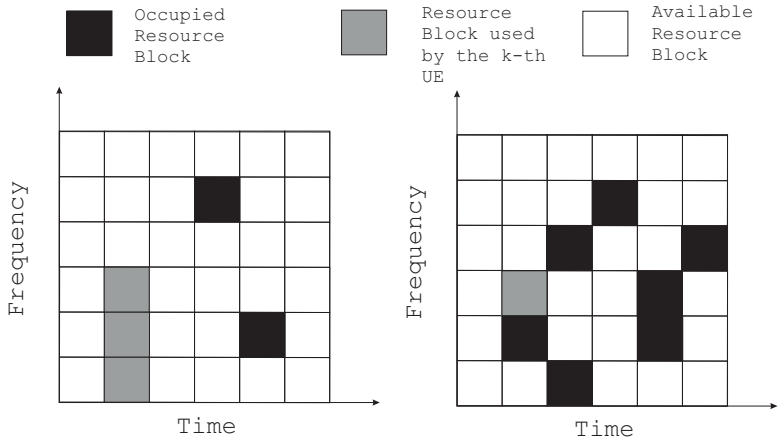


Figure 4.3: Graphic representation of the network under low traffic load (left) and high traffic load (right), respectively. The black, grey and white boxes are used to indicate the occupied time-frequency resource blocks, those used by the intended k-th user and the resources blocks that are available for transmission, respectively.

Generally speaking, the throughput of the system depends on the block error rate (BLER) performance, that accounts for the number of correctly received bits. At the k-th destination node the number of bits received without error can be expressed as follows

$$T_k = m_s^b[(1 - \mathbb{P}_{e,d}) + N_c(1 - \mathbb{P}_{e,r})], \tag{4.16}$$

where $\mathbb{P}_{e,d}$ and $\mathbb{P}_{e,r}$ are the block error probability at the destination and relay node, respectively. Thus, with $N_d \geq 1$ destination nodes, the packet throughput per channel use is given by $\sum_{k=1}^{N_d} T_k$ and measured in bits/s. Assuming that the system is allowed to use more than one resource block and the transmit power per resource block is P/N_{rb}, then the effective packet throughput for the virtual MU-MIMO that uses the N_{rb} available resource blocks is given by

$$E_T = N_{rb} \sum_{k=1}^{N_d} T_k. \tag{4.17}$$

Finally, the average energy efficiency can be defined as

$$EE = \frac{E_T}{P_{\text{cons}}} \qquad \text{[bits/J]}. \tag{4.18}$$

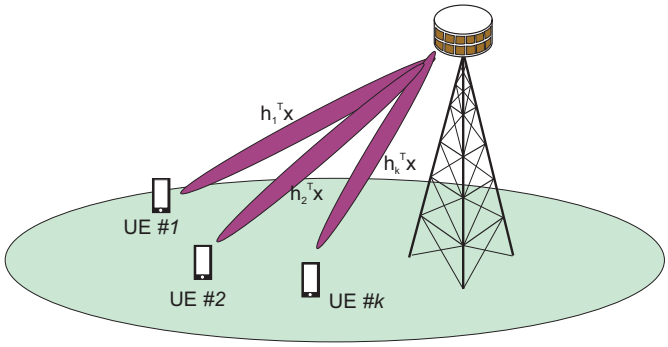


Figure 4.4: Base station simultaneously transmitting to multiple single-antenna users.

4.3.4 Power Allocation

Recall that for large antenna MIMO systems the transmit power increases linearly with the number of transmit antennas at each base station. Thus, power control schemes are used to adapt the transmit power of the base station in order to increase the capacity of the system through interference management or managing the cell coverage [130–133]. Due to inter-user interference, the SINR may fluctuate at the receiver side causing degradation of the system's performance. Power allocation is considered as a suitable way to deal with these detrimental effects, since the transmitted power can be varied to maintain a certain quality of service, minimize inter-user interference or even reduce the total transmit power [134]. Nevertheless, the design of power control schemes is not a trivial task as there are many trade-offs and practical constraints that should be considered. In general, the power allocation optimization problem can be written as

$$\begin{aligned} \max_{v_1,\ldots,v_{N_r}} \quad & g(\text{SINR}_1,\ldots,\text{SINR}_{N_r}) \\ \text{s.t.} \quad & \text{SINR}_k = \frac{|\mathbf{h}_k^T\mathbf{f}_k|^2}{\sigma_N^2+\sum_{i\neq k}|\mathbf{h}_k^T\mathbf{f}_i|^2}\forall k, \\ & \sum_{k=1}^{N_r}\mathbf{f}_k^H\mathbf{f}_k \leq P_{d,k} \end{aligned} \tag{4.19}$$

where $g(\cdot)$ is a performance function dependant on the SINR values, the vector $\mathbf{f}_k$ is the precoding vector at the transmitter for user k and $P_{d,k}$ is the power allocation per user at the transmitter.

The problem in (4.19) is in general complex to solve, since there might be (a) conflicting in-

terests between users, (b) strong inter-user interference, (c) a non-convex performance region and (d) a large set of feasible beamforming vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_{N_r}\}$. Nevertheless, ZF precoding removes interference by projecting the observations as

$$\mathbf{v}_i = \sum_{k=1}^{N_r} (\mathbf{h}_k \mathbf{h}_k^H)^* \mathbf{h}_i, \tag{4.20}$$

which is the orthogonal complement of the interfering signals. This condition greatly simplifies the precoding design by reducing the search-space and is of practical importance in high-SNR scenarios where inter-user interference dominates the noise term in the SINR expression [135]. Although perfect interference nulling requires perfect CSI, ZF precoding can be implemented under imperfect CSI by avoiding inter-user interference along some of the strongest eigenvectors $\mathbb{E}\left[\mathbf{h}_k \mathbf{h}_k^H\right]$ of each user k [135–137].

The power allocation, $P_{d,k}$, that will determine the operating point in the performance region can be computed explicitly for ZF precoding, since all the effective interfering channels are zero. For a set of given power coefficients $P_{d,k} > 0$ and invertible performance functions $g_k(\cdot)$, the power allocation problem can be described as [138]

$$\begin{aligned} &\max_{P_{d,k}>0} \sum_{k=1}^{N_r} w_k g_k \left(P_{d,k} \frac{\rho_{k,k}}{\sigma_N^2} \right) \\ &\text{s.t.} \sum P_{d,k} \leq P^{\max} \end{aligned} \tag{4.21}$$

where $\rho_{i,k} = |\mathbf{h}_k^T \mathbf{f}_i|^2$, $P^{\max}$ is the maximum transmit power and w_k is an associated cost function weight. The solution of (4.21) is given in [138] as follows:

$$P_{d,k} = \left[\frac{\sigma_N^2}{\rho_{k,k}} \frac{d}{dx} g_k^{-1} \left(\frac{\sigma_N^2}{f_k w_k \rho_{k,k}} \right) \right]_+, \tag{4.22}$$

where the function $[\cdot]_+$ replaces negative values with zero, and the parameter $f_k \geq 0$ is selected to use full power.

4.4 Numerical Results

In this section, the energy efficiency analysis of the virtual MU-MIMO system described in Section 4.2 is presented. A constraint on the number of available resource blocks and the available information used for channel estimation is considered. Unless stated otherwise, the results shown in this section pertain to a BICM-MIMO system with M-QAM Gray mapping modulation and a Rate-1/2 convolutional code with constraint length 7 and generator polynomials $[133, 171]_{\text{octal}}$. Rayleigh block-fading with 10^5 blocks, where each block has 200 consecutive symbol periods, is considered. Moreover, ZF linear precoding is performed before transmission and the decode-and-forward protocol is implemented at each relay node to enable wireless cooperation. Further, the total transmitted power P is normalized to unity. The simulation results are computed through the Monte Carlo method. Other simulation parameters are given in Table 4.1 and were adopted from [7, 55, 123].

Description	Parameter	Value
Power allocation ratio	γ	0.25
Efficiency of the PA at BS	δ	0.3
Fixed power consumption	P_0	18 W
Power consumed by oscillator	P_{lo}	2 W
Circuit power at BS	P_{ct}	1 W
Circuit power at UE	P_{cr}	0.1 W
Power required for coding	P_{cod}	0.1 W/(Gbit/symbol)
Power required for decoding	P_{dec}	0.8 W/(Gbit/symbol)

Table 4.1: Simulation Parameters of Section 4.4.

4.4.1 Perfect Channel State Information

In order to understand the results presented in this section, Figure 4.5 shows the power consumption model described by equation (4.12). As can been seen, the model is linearly dependent on the number of transmit antennas and the transmit power consider at the base station. Further, equation (4.14) was used to compute the spectral efficiency, SE $= W \times R_k$, of a MIMO system with ZF precoding. From Figure 4.6, it can be seen that as the number of users increases for a constant number of transmit antennas, $N_t = 20$, the spectral efficiency follows a quadratic function. In other words, after reaching an optimal number of UEs ($N_r = 10$) served by the BS, the reduction of the degrees of freedom in ZF beamforming scales down the achievable rate. In the following figures, the effect that the wireless cooperation carried out at the receiver

has on the energy efficiency performance is studied.

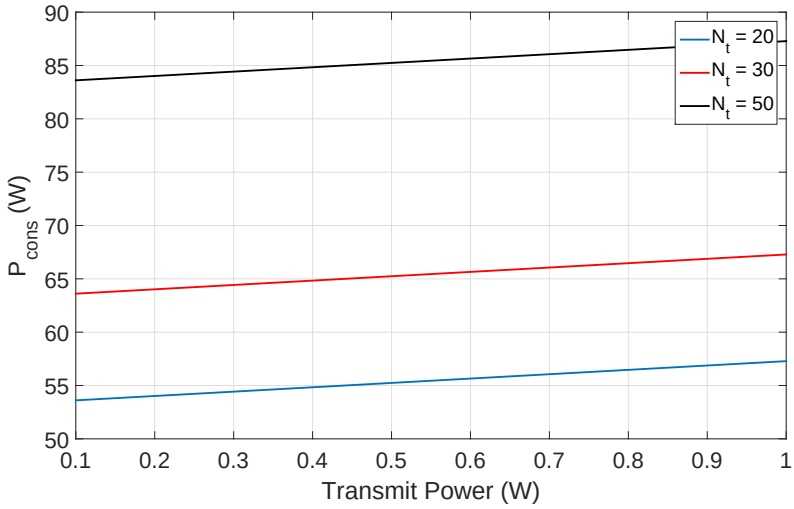


Figure 4.5: Power model described by equation (4.12) as a function of the number of transmit antennas and maximum transmit power at the base station.

Figure 4.7 compares the effective packet throughput when the target rate has been set as 8 bits/s and $N_d = 1$ and $N_t = 20$ remain constant. It can be seen that the optimal scenario in terms of effective throughput is obtained when the system uses as many resource blocks as possible (Case 1). For this particular scenario, the degrees of freedom are higher compared with the other cases improving the global performance. Moreover, there is no overhead power consumed since there are no active relays. From Figure 4.7, it can also be noticed that the power required to achieve the target rate increases as the modulation order does.

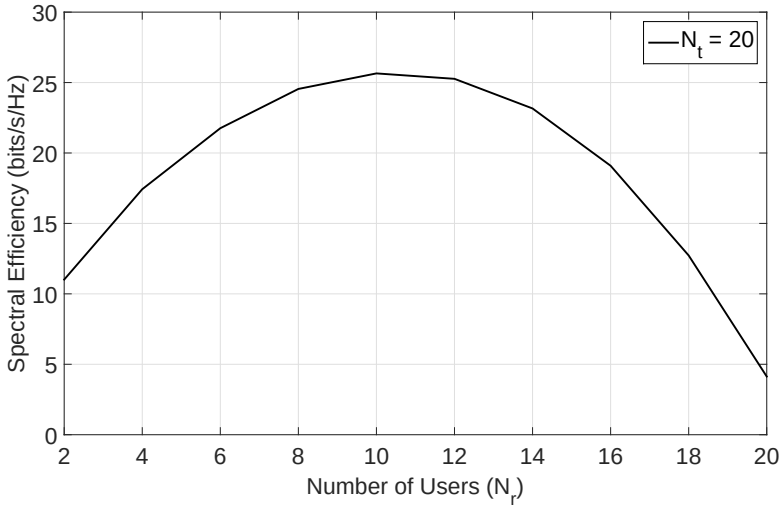
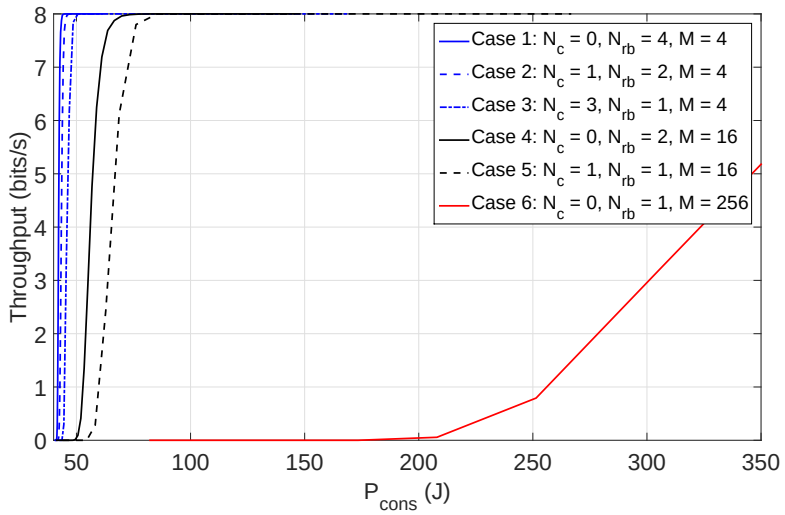
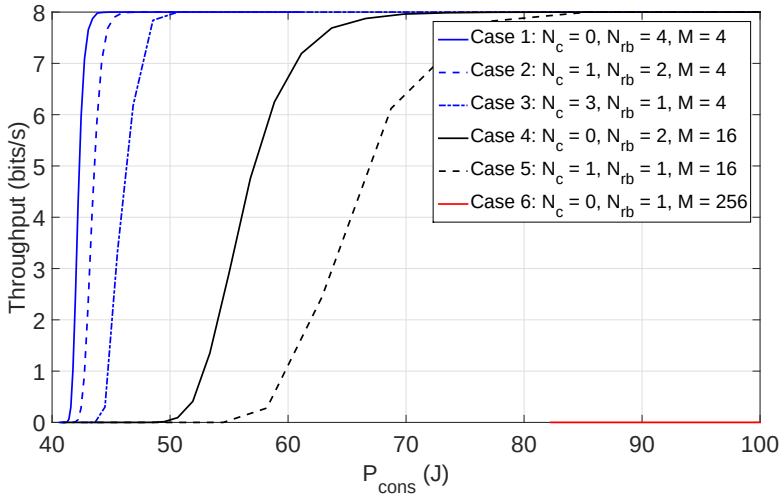


Figure 4.6: Achievable rate of a MIMO system that implements ZF precoding with $N_t = 20$.

As mentioned, it is of particular interest to analyse the energy consumption of V-MIMO systems



(a)



(b)

Figure 4.7: Packet throughput of a virtual MIMO system with $N_t = 20$, $N_d = 1$ and a target rate of 8 bits/s, where N_c is the number of relays and N_{rb} is the number of resource blocks used for transmission and M is the modulation order. Figure 4.7(b) shows a zoomed version of Figure 4.7(a).

with constrained resource blocks, thus Figure 4.8 shows the EE for the aforementioned settings regarding the effective packet throughput. In order to calculate the value of EE, the transmit power of the system was increased until the target packet throughput was reached. From Figure 4.8, it can be seen that the EE grows with E_T until the system reaches the target rate of 8 bits/s. These results are related with the behaviour observed in Figure 4.7, where the throughput increases with the transmit power. However, once the maximum packet throughput delivered

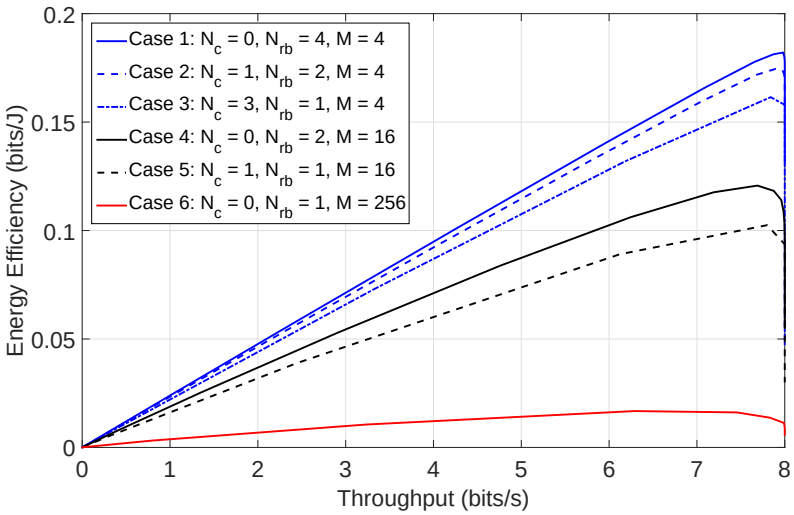


Figure 4.8: Energy efficiency of a virtual MIMO system with $N_t = 20$, $N_d = 1$ and a target rate of 8 bits/s, where N_c is the number of relays and N_{rb} is the number of resource blocks used for transmission and M is the modulation order.

by the system is reached, increasing the transmit power does not affect the system in terms of throughput, but it does affect the EE. Since the BS is consuming more power to achieve the same throughput, then the EE drops as can be seen in Figure 4.8. Knowing the point where the system achieves the target throughput is important to maintain a high EE and to save energy. Finally, it can be noticed that the highest EE is obtained when the system uses all the available resource blocks. However, a good and practical trade-off between EE and the required resource blocks is shown for the Cases 2 and 3. In particular, Case 3 shows a comparable EE performance using a DF cooperation between users and only one resource block. Therefore, it can be considered that the cooperative system is a viable option for a practical systems where the number of available resource blocks is constrained.

Continuing with the analysis of the E_T and EE, it is now considered that the system is limited to $N_{rb} = 1$ and there is more that one destination node to be served. The target rate of the system is set to 16 bits/s, the number of transmit antennas is $N_t = 20$ and a variable number of relays, N_c, per destination node is considered. From Figure 4.9 and comparing Cases 7-9, it is possible to notice the contribution of the relays to the total power consumption. Different from Case 9, where only one relay is assisting each destination node, the increment in the overhead power consumption is more evident in Case 7 where each destination node is assisted by 7 relays. Recall from Section 4.3.1, that this overhead is result of the power consumed for decoding, re-encoding and forwarding the symbol at the relays. From this plot, it can also be noted that

the system requires less power to reach the target throughput when the number of beams is increased and not the modulation order.

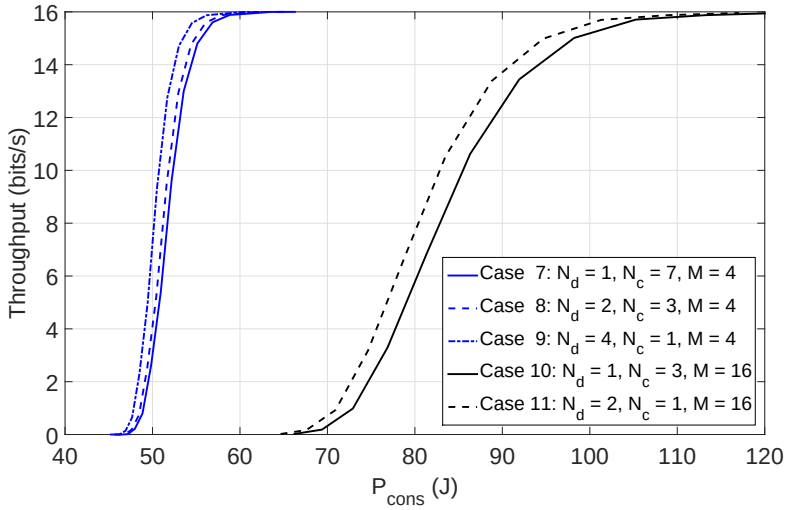


Figure 4.9: Packet throughput of a virtual MIMO system with $N_t = 20$, $N_{rb} = 1$ and a target rate of 16 bits/s, where N_c and N_d are the number of relays and destination nodes, respectively, and M denotes the modulation order.

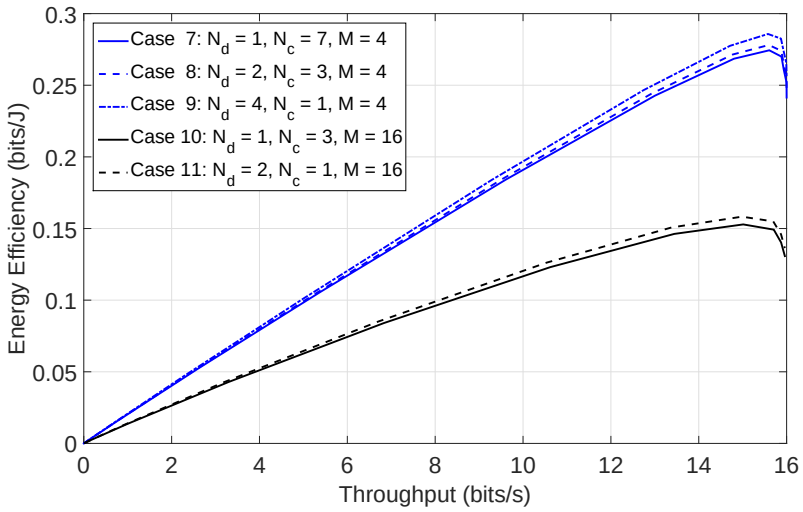


Figure 4.10: Energy efficiency of a virtual MIMO system with $N_t = 20$, $N_{rb} = 1$ and a target rate of 16 bits/s, where N_c and N_d are the number of relays and destination nodes, respectively, and M denotes the modulation order.

Finally, Figure 4.10 compares the EE performance when more that one destination node is served by the BS. As can be seen, when the number of relays per destination node goes from $N_c = 1$ to $N_c = 7$ (Case 7 and 9) with $M = 4$, the EE achieves a maximum value of 0.28 and 0.27 (measured in bits/J), respectively. This reduction in the EE is result of the increase of total power consumption (see Figure 4.9). Further, as shown in Figure 4.10, when $M = 16$ the

maximum EE achievable is 0.15 bits/J. Thus, for a fixed throughput the EE diminishes when the modulation order increases.

4.4.2 Imperfect Channel State Information

In this subsection, the impact of the imperfect channel estimation in the performance of the network is investigated. In order to simulate a system with imperfect CSI at the transmitter, the channel was modelled by using equation (4.2). The analysis is particularly focused on the energy efficiency and the trade-off between power consumption, throughput and channel estimation error. Figure 4.11 presents the throughput achieved by the virtual MIMO system when $N_r = 4$ ($N_d = 1$, $N_c = 3$) and a 4-QAM modulation order is used per antenna. As can be

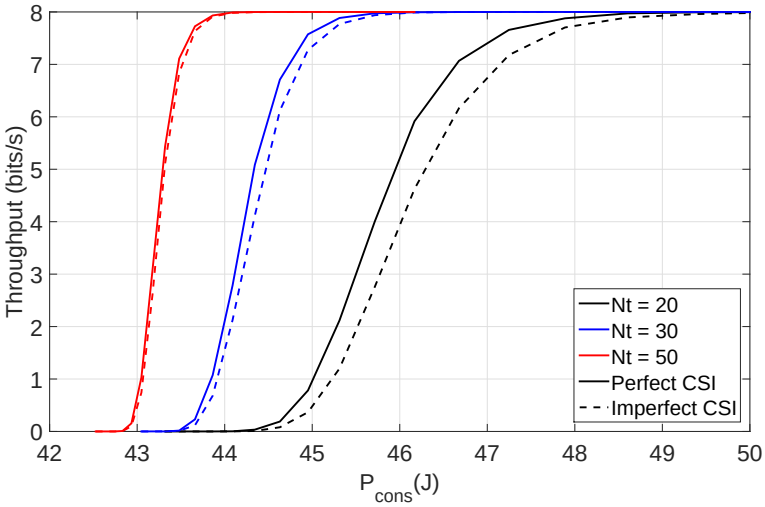


Figure 4.11: Packet throughput of a virtual MIMO system with $N_t = 20$, $N_d = 1$, $N_c = 3$ and a target rate of 8 bits/s when equal transmit power is used per antenna.

seen, as the number of transmit antennas increases from $N_t = 20$ to 30 and 50, the performance gap between the system with perfect CSI and imperfect CSI measured at 4 bits/s is reduced from 0.5 J to 0.25 and 0.10 J, respectively. From this plot, it can also be pointed out that, even though no power allocation scheme was employed, the degrees of freedom provided by increasing the number of transmit antennas mitigated the effects of the imperfect CSI. Moreover, the system's performance in terms of energy efficiency is presented in Figure 4.12. As pointed out in the previous subsection, as the number of antennas increases so does the total power consumption, nevertheless the improvement in terms of throughput is considerable. This means that the system with $N_t = 50$ transmit antennas has the best performance in terms of energy efficiency. Moreover, it can be seen that the performance gap for the maximum achievable EE value, when

both perfect and imperfect CSI are considered, reduces as the number of transmit antennas increases as a result of the trade-off between throughput and power consumption imposed by the system.

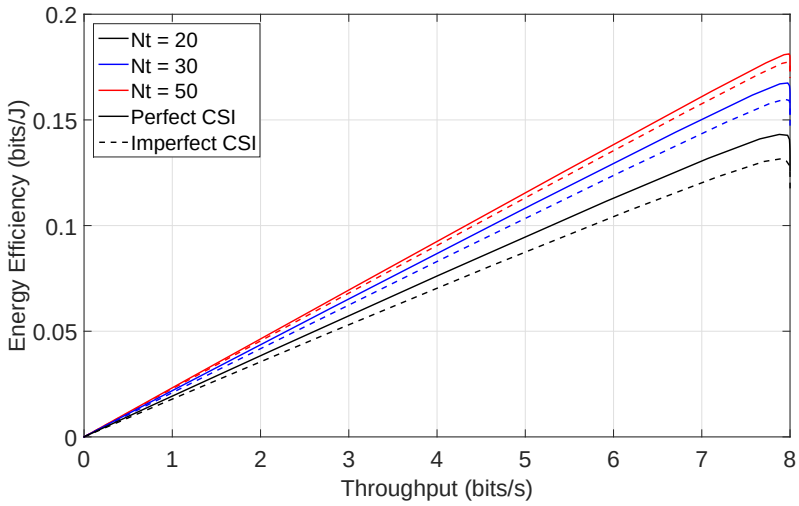


Figure 4.12: Energy efficiency of a virtual MIMO system with $N_t = 20$, $N_d = 1$, $N_c = 3$ and a target rate of 8 bits/s when equal transmit power is used per antenna.

4.4.3 Power Allocation

As it was mentioned, the transmit power of large antenna MIMO systems increases linearly with the number of transmit antennas. Thus, power control schemes can be used to adapt the transmit power of the base station in order to increase the capacity and reduce the total power consumption. In this subsection, the system's performance in terms of energy efficiency and throughput, when the power allocation scheme described in Section 4.3.4 is employed, is evaluated. For illustrative purposes, the power coefficients $P_{d,k}$ obtained by using equation (4.22) for a virtual MIMO system equipped with $N_t = 4$ transmit antennas are given by $\{0.5958, 0.6509, 0.3806, 0.2768\}$. Thus, instead of assuming a equal power level for transmission, the power allocation algorithm uses the coefficients that improve the system's throughput.

Continuing with the analysis and similar to the previous section, the virtual MIMO system is assumed to be equipped with $N_r = 4$ ($N_d = 1$, $N_c = 3$) receivers and 4-QAM modulation. As can be seen from Figure 4.13, the larger the number of transmit antennas, the lower is the power consumption required to achieve the target packet throughput, this result is mainly associated with the fact that the precoding matrix possesses more degrees of freedom for large antenna MIMO systems. From this plot, it can also be pointed out that compared with the case where

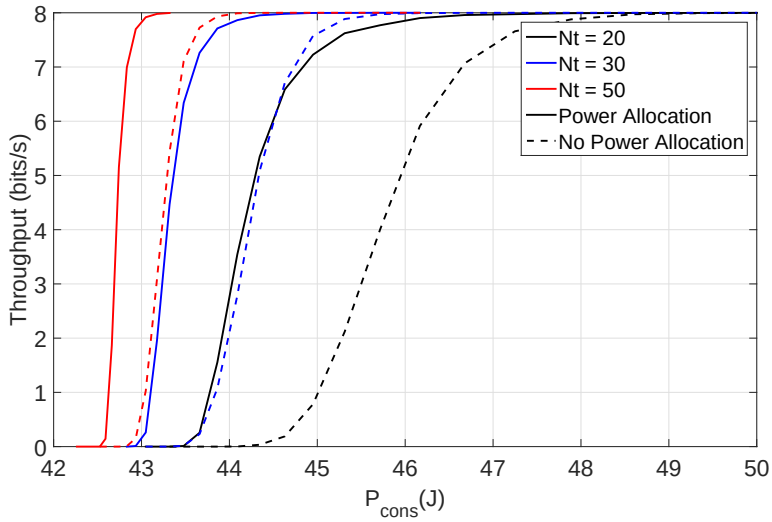


Figure 4.13: Packet throughput of a virtual MIMO system with $N_t = 20$, $N_d = 1$ and a target rate of 8 bits/s when the power allocation algorithm described by equation (4.22) is used.

the maximum transmit power is used per antenna, when the power allocation algorithm is used, it is possible not only to reduce the total power consumption of the system but also to use the transmit power in an efficient way. This provides higher gains to the links that are more affected by the Rayleigh fading. In particular, the gap between the systems with and without the power allocation algorithm at target packet throughput of 4 bits/s and $N_t = 20$, 30 and 50 are 1.57, 0.92 and 0.52 J, respectively. Notice that the performance gap reduces as the number of transmit antennas increases, this can be explained by considering that with a relatively low number of transmit antennas the opportunity for improvement is considerable, while very large MIMO systems mainly take advantage on the fact that the effects of fading, imperfect CSI and interference vanish as the number of transmit antennas grows [139].

Considering the performance in terms of packet throughput, Figure 4.14 compares the energy efficiency performance of the virtual MIMO system when power allocation is implemented. From this plot is evident that not only the larger the number of transmit antennas the better the performance that can be achieved but also that the maximum performance gap between the EE values decreases as the number of transmit antennas increases, going from 0.014 to 0.009 and 0.007 when $N_t = 20$, 30, 50, respectively. These results are clearly associated with those shown in Figure 4.13 were the power required to carried out the transmission and cooperation can lower when a power allocation algorithm is used compared to the case where a maximum transmit power per antenna is employed for transmission.

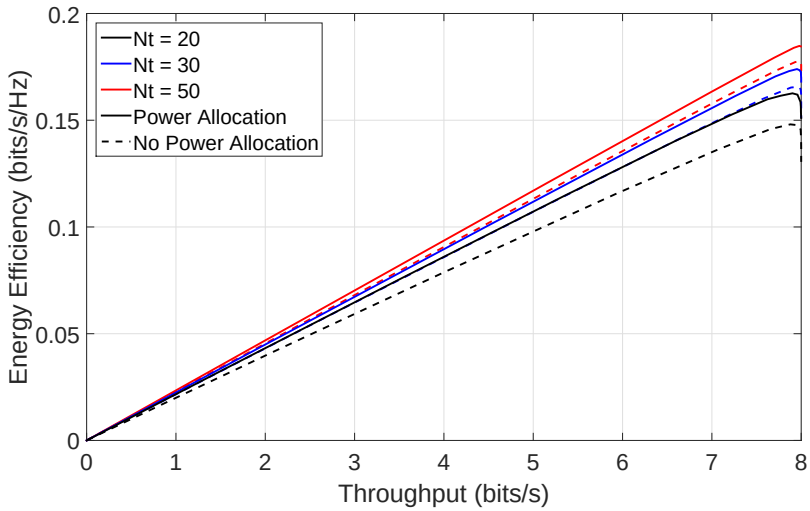


Figure 4.14: Energy efficiency of a virtual MIMO system with $N_t = 20$, $N_d = 1$ and a target rate of 8 bits/s when the power allocation algorithm described by equation (4.22) is used.

4.5 Summary

In this chapter, a cooperative virtual MU-MIMO system that uses a multi-antenna transmitter and implements BICM transmission and DF cooperation among closely spaced single-antenna receivers was presented. Large multi-antenna systems provide an effective way to enhance the spectral efficiency of wireless networks at the expense of linearly increasing circuit power consumption due to the use of multiple radio frequency chains. In this chapter, the energy efficiency and the effective throughput of this wireless cooperative system were analysed. In order to provide a common framework, a power consumption model that considers not only the transmit and circuit power consumed at the transmitter but also that takes into account the overhead power consumption incurred by the wireless cooperation was introduced. Then, the throughput and energy efficiency of the cooperative system with both perfect and imperfect channel state information were evaluated.

When perfect channel estate information is considered, it was shown that the best performance in terms of throughput and energy efficiency is obtained when the number of available resource blocks for transmission is not restricted. Nevertheless, when a practical system was considered, i.e., the number of resource blocks is constrained, virtual MIMO showed to be an enabler for future wireless communication systems. It can be highlighted that when the target throughput of the system was fixed, the maximum energy efficiency value achieved for $M = 4$ was 0.2 bits/J. Meanwhile when $M = 16$ was employed, the maximum EE obtained was 0.15 bits/J.

This gap was result of the increase of the total power consumption required to meet the target rate with different modulation orders, going from 60 to 110 J.

Moreover, simulation results showed that the larger the number of transmit antennas, the lower the impact of the imperfect CSI on the throughput and energy efficiency of the system, where a performance gap between the system with perfect CSI and imperfect CSI measured at 4 bits/s was reduced from 0.5 J to 0.25 and 0.10 J, when $N_t = 20$, 30 and 50, respectively. Finally, when a power allocation algorithm was considered, the gap between the systems with and without power allocation at 4 bits/s and $N_t = 20$, 30 and 50 was 1.57, 1.02 and 0.52 J, respectively. Thus, it was pointed out that the performance gap in terms of throughput reduces as the number of transmit antennas increases. Meanwhile, the maximum EE value was obtained for $N_t = 50$, the maximum number of transmit antennas considered in this analysis.

In the next chapter, the energy efficiency of a self-backhauling system is evaluated and sleeping strategies are used to improve the energy efficiency of cellular systems with inhomogeneous distribution of the users.

Chapter 5
Energy Efficient Inhomogeneous Cellular Networks

In this chapter, a wireless self-backhauling heterogeneous network is considered. Each micro base station is assumed to be equipped with large multi-antenna transmitter to perform backhaul and access operations using mm-wave frequencies. Furthermore, a decode-and-forward (DF) cooperation algorithm is employed to provide a cost-effective and reliable backhaul links between micro BSs.

An energy efficient base station deployment and operation by using the heuristic greedy algorithm are presented. Due to the inhomogeneity of the traffic load in 5G heterogeneous networks, first, the users are assumed to be inhomogeneously distributed in space with the average number of users varying over time. Later, a fixed average number of users is assumed to be served by each base station while their distribution in space is modelled by a Thomas Cluster Process.

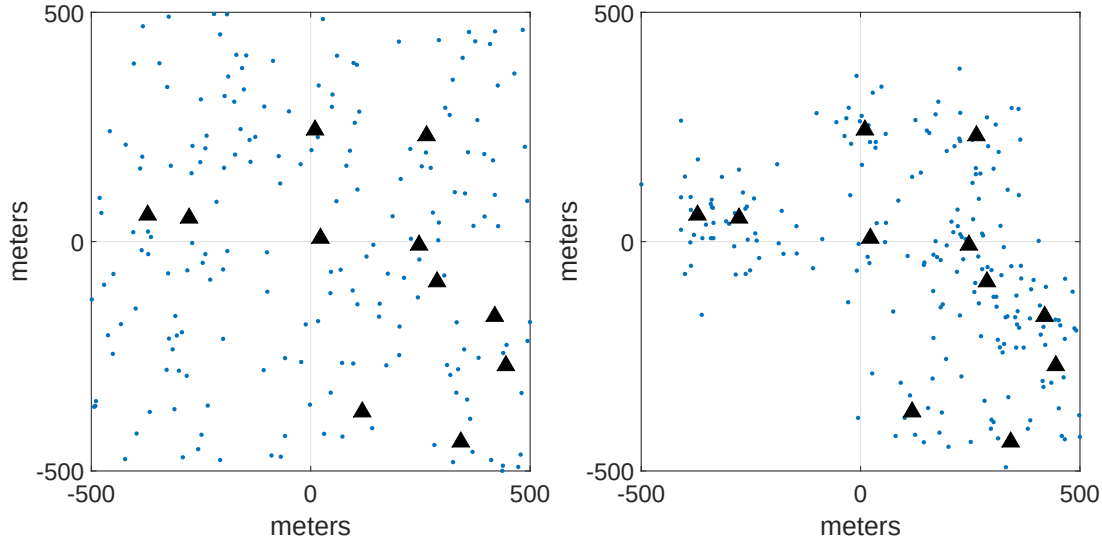
Numerical results show that by using the greedy algorithm for base station deployment, the average number of micro base stations required to support the traffic load at peak-time can be reduced. On the other hand, highly clustered networks exhibit a smaller average serving distance and thus a better probability of coverage. Hence, the greedy algorithm provides a higher throughput than the load-based and random algorithms by considering the limitations in frequency resources for backhauling while switching-off base stations.

The remainder of this chapter is organised as follows. Section 5.1 presents the background, motivation and problem definition of this chapter. The greedy algorithm, that is used throughout the chapter, is introduced in Section 5.2. An efficient base station deployment for heterogeneous networks is presented in Section 5.3. A description of the sleeping strategies and their corresponding performance analysis in terms of coverage and energy efficiency are presented in Section 5.4. Simulation results for the base station deployment and sleeping strategies are shown in Section 5.5. Finally, Section 5.6 concludes the chapter.

5.1 Introduction

Heterogeneous networks (HetNets) have been considered as a promising system architecture to meet the exponentially increasing demand for wireless data traffic [71, 140–142]. In HetNets, the macro-cell BSs control the user scheduling, resource allocation and support high-mobility users, while micro BSs, with a smaller coverage area, provide high data rates for low-mobility users. In this scenario, not only can a better frequency reuse can be achieved, but also the energy efficiency (EE) of the network can be improved due to reduced path loss in small cells [143]. Due to the large traffic load that is passed into the core network, one of the major impediments to deploying of HetNets is access to cost-effective, reliable and scalable backhaul networks [144]. Since the traditional optical fibre backhaul is not always available due to high deployment costs, the use of millimetre-wave (mm-wave) bands has been proposed to enable broadband radio and backhauling in future wireless networks [145]. The advantages of mm-wave backhaul include overcoming geographical constraints, the availability of large bandwidth, easy deployment of multi-antenna transmission and typical line-of-sight (LoS) propagation characteristics that help to control co-channel interference [140, 144, 145]. Nevertheless, in order to compensate for the high path loss large array and beamforming gains are required [144]. Since relaying allows the system to reduce signal degradation due to path loss by shortening the distance between transmitter and receiver, cooperative systems can be used to relay the backhaul data from the micro BSs to the aggregation points, which are BSs with wired backhaul. A well-researched method to model HetNets is the assumption of independent homogeneous Poisson point processes (PPP) to determine both users' and base stations' locations. Despite the analytical tractability this approach lacks accuracy to model future 5G HetNets, since the micro BSs are to be deployed in locations where the users are more likely to appear, so the UEs will be clustered towards the BS, rather than being uniformly distributed [146, 147]. Furthermore, the frequent assumption of unlimited capacity for backhauling is unrealistic. Among the targets of the 5G mobile and wireless communications systems is to handle the inhomogeneous distribution of increasing traffic demand over time and space in a efficient manner [146].

In order to capture the dynamics of 5G HetNets, the spatial distribution of the users within the area of interest (see Figure 5.1) can be viewed as an additional degree of freedom to the wireless network. These networks are frequently referred to as heterogeneous HetNets (HetHetNets) since they exhibit heterogeneity in both infrastructure and user distribution [148]. The user distribution is a critical parameter that will determine the network performance in wireless cellular



(a) (b)

Figure 5.1: Comparison of the (a) homogeneous PPP and (b) clustered distribution of the users within a 1000m $\times$1000m area of interest, where the triangles and dots represent the BSs' and UEs' locations, respectively.

networks. There are two key performance metrics that are affected by the user distribution: (1) the load of the BSs which is determined by the clustering properties of the user distribution, and (2) the spectral efficiency of the UEs which is a function of signal-to-interference-plus-noise ratio (SINR) and the bandwidth [141, 147].

5.1.1 Related Work

In [149], the authors study the spatial traffic heterogeneity of outdoor users in dense areas of a urban scenario via a second-order statistic: the Coefficient of Variation (CoV) of two spatial metrics of the resulting point process, the Voronoi cell areas and the Delaunay cell edge lengths. The results show that the topology of the buildings in a urban environment imposes a significant degree of heterogeneity on the spatial distribution of the wireless traffic. In [150], a tractable method of sampling UEs by conditionally thinning the BS point process shows that the resulting framework can be used as a generative model to study current capacity-centric deployments, where the UEs are more likely to lie closer to the BSs. Moreover, in [151], closed-form analytical models for signal-to-interference-ratio (SIR) and coverage in cellular networks with heterogeneous traffic are presented when macro-BSs with omnidirectional antennas are considered.

An adjustable spatial traffic model for HetHetNets with limited backhaul connection capacity is introduced in [141]. It is shown that user-in-the-loop (UIL) schemes, whereby the BS load and backhaul connections are taken into account, improve the mean user rate and the coverage probabilities in future 5G networks. Nevertheless, all the backhaul connections are assumed to be wired. Meanwhile, the authors in [148] introduce a spatial traffic modelling approach with adjustable statistical properties capturing the severity of the network heterogeneity by using the coefficient of variation and the correlation coefficient between the spatial UE and BS distributions. In [147], an analysis of the downlink coverage probability is presented. The results are specialized to a particular case of a Thomas cluster process demonstrating that the coverage probability decreases when the variance of the user location distribution increases, ultimately collapsing to the result for the PPP user distribution. In [152], the authors provide a formal analysis of the downlink coverage probability in terms of a general density function describing the locations of users around the BSs. The derived results are specialized for two cases of interest: the Thomas cluster process and the Matérn cluster process.

Furthermore, in [153], the authors proposed a greedy algorithm that finds the set of micro BSs locations that maximizes the EE of the network while satisfying the capacity demand, though, multiple antenna transmission is not considered in this approach. In order to minimize the total energy consumption of a HetNet, in [154], the authors decompose the optimization into a deployment problem at peak-time and an operation problem at off-peak time. In other words, the network was designed to satisfy the quality-of-service of a target traffic load and optimized to operate when the traffic load falls below this threshold. For the deployment optimization, they proposed a constant-factor approximation greedy algorithm, while the dynamic operation problem uses a Lagrangian relaxation solution. In [155], the authors analysed the EE and its relationship with the densities of macro and micro BSs in a HetNet. In particular, they used a stochastic geometry based model to derive the EE and area power consumption. Meanwhile, the area spectral efficiency (ASE) aspect of micro BS deployment is studied in [156]. Furthermore, aiming to save energy depending on the traffic profile of the cell, switching-off algorithms for macro and micro BSs during off-peak times are investigated in [143] and [156].

5.1.2 Problem Definition

The following sections are aimed to improve the energy efficiency performance for the self-backhauling heterogeneous network. Without loss of generality, the optimization problem for

the EE can be formulated as follows

$$\begin{aligned} &\min_{||r||,P} \sum_{b\in\mathcal{B}} \eta_{EE}(b) \\ \text{subject to} \quad &P_{se} \leq P_{cons} \\ &f_w R_{bh} \geq R_{fh}(1-f_w) \end{aligned} \tag{5.1}$$

where $\eta_{EE}(b)$ is the sum EE value that can be achieved by the set of BSs, $\mathcal{B}$, the transmit power is denoted by P, $||r||$ is the serving distance and P_{se} is the power consumption of the network once some BSs have been switched-off. If f_w represents the fraction of resources used for backhauling, an in-band solution is only feasible if $f_w R_{bh} \geq R_{fh}(1-f_w)$, where R_{fh} denotes the sum data rate of the access links and R_{bh} is the sum data rate of the backhaul links supporting the network.

There are many reasons why equation (5.1) may be difficult to solve in an optimal fashion. The most important might be: (a) conflicting interests between energy efficiency, throughput and power consumption; (b) non-negligible inter-channel interference; (c) the performance region can be non-convex [142, 157] and (d) the computational complexity of the combinatorial problem grows exponentially with the number of base stations within the network [154].

The contribution of this chapter can be summarised as follows:

(a) A heuristic greedy algorithm for energy-efficient BS deployment and dynamic BS operation is proposed.

(b) A tunable model that can cover a wide range of scenarios from a homogeneous PPP to highly heterogeneous clustered wireless networks is presented. A *Thomas cluster process* is employed to generate the clustered UEs.

(c) An analysis of the probability of coverage and throughput is provided for a wireless self-backhaul system with constrained frequency resources.

(d) An energy saving assessment is performed by using random, load-based and greedy sleeping policies.

In the following, the principles of the greedy algorithm are introduced in Section 5.2 and the optimization problem in (5.1) is analysed from two different perspectives.

5.2 Greedy Algorithm

Recently, the study of combinatorial problems with submodular objective functions has attracted great attention, since many optimization problems can be represented as constrained variants of submodular maximization [158]. In general, the techniques used to compute approximate solutions to submodular maximization problems can be divided into two main categories. The first approach is combinatorial in nature, and is mostly based on local search techniques and greedy rules. In this approach the solutions are usually tailored for specific structures making extensions quite difficult. On the other hand, the second approach is composed of two steps. In the first step, a fractional solution is found for a relaxation of the problem. In the second step, the fractional solution is rounded to obtain an integral one while incurring only a small loss in the objective. Two issues arise when using this approach. First, since the objective function is not linear, it is not clear how to formulate a relaxation which can be approximated efficiently. Second, given a fractional solution, a rounding procedure that outputs an integral solution without losing too much in the objective function is required. In this work, the greedy approach is used to approximate the optimization problem presented in (5.1).

A greedy algorithm is a mathematical process that looks for simple, easy-to-implement solutions to complex, multi-step problems by making the choice that at each step provides the most obvious benefit [159]. These algorithms are called greedy because even though the optimal solution to each smaller instance will provide an immediate output, the algorithm does not consider the larger problem as a whole, i.e., it makes a locally optimal choice in the hope that this choice will lead overall to a globally optimal solution. Once a decision has been made, it is never reconsidered.

Greedy algorithms work by recursively constructing a set of objects from the smallest possible constituent parts. Amongst the advantages of using a greedy algorithm, it is possible to highlight the fact that solutions to smaller instances of the problem can be straightforward and easy to understand. The disadvantage is that a series of local optimal solutions may not lead to the global optimal solution. Nevertheless, the greedy algorithm guarantees a reasonably good approximation to the optimal solution if the function f is submodular. Assuming that $f(\emptyset) = 0$, i.e., the empty set carries no value, then the submodularity property has two equivalent definitions described as follows [160]

Definition 5.1 (Submodularity)**.** *Let X be a finite set, commonly called the ground set. A function $f : 2^X \rightarrow \mathbb{R}$ that assigns each subset $S \subseteq X$ a value $f(S)$ is called submodular if for*

all subsets $S \subseteq T \subseteq X$ *and all* $x \in X \setminus T$,

$$f(S \cup \{x\}) - f(S) \geq f(T \cup \{x\}) - f(T). \tag{5.2}$$

In other words, if f measures the *"benefit"*, then the marginal benefit of adding x to S is at least as high as the marginal benefit of adding it to T. Since $S \subset T$ and x are all arbitrary, then, adding x to a bigger set cannot provide a better *benefit* than adding it to a smaller set. Equivalently,

Proposition 5.1. *f is a submodular function if and only if for all $S, T \subset X$, it holds that*

$$f(S \cap T) + f(S \cup T) \leq f(S) + f(T). \tag{5.3}$$

Therefore, submodular set functions exhibit a natural diminishing returns property [160]. A property of submodular functions is that the non-negative linear combination of submodular functions is also a submodular function. In other words, if $f_1, \ldots, f_k$ are submodular on the same set X, and $\alpha_1, \ldots, \alpha_k$, with $\alpha_k \in \mathbb{R}^+$, then $\alpha_1 f_1 + \cdots + \alpha_k f_k$ is also a submodular function on X [160, 161].

Another property that is required on a submodular function is monotonicity. In other words, the more elements in the ground set, the larger the value under f. Thus, f is a monotonic function if for every $S \subseteq T$ then $f(S) \leq f(T)$. A useful property of functions that are both submodular and monotonic is that the truncation of such a function is also submodular and monotonic. In other words, the function $\min(f(S), c)$ is still a submodular if f is submodular and monotonic and $c \in \mathbb{R}$ is a constant.

Definition 5.2. *Let f be a monotonic, submodular, non-negative function on X. The greedy algorithm, which starts with S as the empty set and at every step selects an element x which maximizes the marginal benefit $f(S \cup \{x\}) - f(S)$, provides as an output a set S that achieves a $(1 - 1/e^1)$-approximation of the optimum.*

The proof of Definition 5.2 can be found in [160, 162]. In the following, some common examples of submodular functions are illustrated [163].

- *Coverage in set.* Let $T_1, T_2, \ldots, T_n$ be subsets of a finite set T. If $S = [n] = \{1, 2, ..., n\}$

is the ground set, then the coverage function $f : 2^S \rightarrow \mathbb{R}_+$ defined as $f(A) = |\cup_{i \in A} T_i|$ is submodular, non-negative and monotonic.

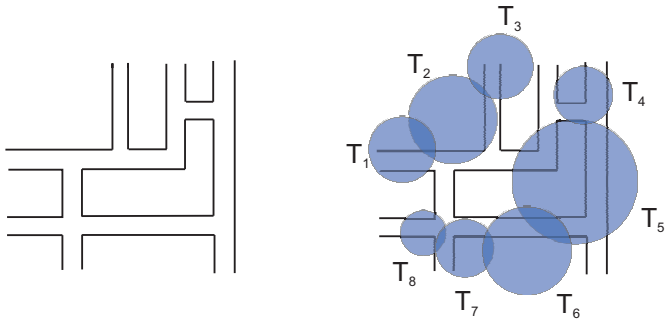


Figure 5.2: Coverage in set system.

- *Combinatorial optimization problems.* Let S be a finite set with n elements and 2^S its power set of cardinality 2^n. If $f : 2^S \rightarrow \mathbb{R}$, then $\min_{A \subset S} f(A)$ is submodular.
- *Entropy and Mutual information.* Let $X_1, X_2, \ldots, X_n$ be random variables, and $S = \{1, 2, ..., n\}$. If $X_A = \{X_i | i \in A\}$ is the set of random variables with indices in A and $A \subseteq S$. Then $f(A) = H(X_A)$, where $H(\cdot)$ is the entropy function, is submodular, non-negative and monotonic. Additionally, $f(A) = I(X_A; X_{S|A})$, where $I(\cdot;\cdot)$ is the mutual information of two random variables, is also submodular.

5.3 Base Station Deployment

In this section, an energy-efficient BS deployment algorithm considering that the users follow an inhomogeneous distribution in space and time is presented. Moreover, an energy-efficient dynamic operation for the additional micro BSs during the off-peak traffic load periods is discussed.

5.3.1 System Model

Consider the downlink of a heterogeneous cellular network where a set of macro and micro base stations, $\mathcal{B}_\phi = \mathcal{B}_M \cup \mathcal{B}$, are assumed to transmit at the power P_M and P_a, respectively (see Figure 5.3). Throughout the chapter, the subscript M is used to denote a macro base station. Moreover, the UEs are associated with the BS that provides the highest signal strength at the location of the user.

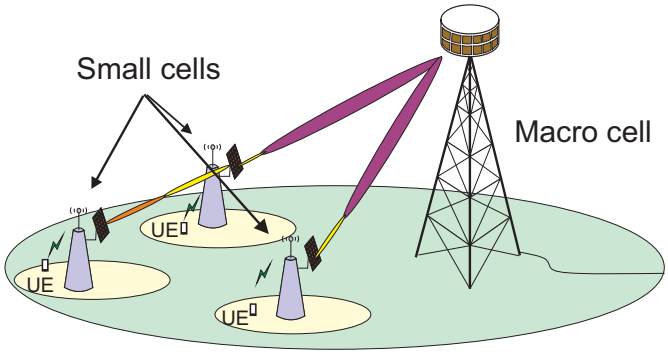


Figure 5.3: Millimetre-wave MIMO based wireless backhaul for 5G.

In this section, it is assumed that a N_t-antenna micro BS communicates with a N_r-antenna micro BS via N_s data streams, where $N_s \leq N_t$ and $N_s \leq N_r$. Moreover, a $N_t \times N_s$ precoder $\mathbf{F}$ is applied at the transmitting BS, with $\| \mathbf{F} \|_F^2 = N_s$. At the destination, a combiner is used to process the received signal which results in

$$\mathbf{y} = \sqrt{P_i}\mathbf{W}^H\mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{W}^H(\mathbf{n} + \mathbf{I_n}), \tag{5.4}$$

where P_i denotes the average transmit power with $i \in \{a, M\}$, $\mathbf{H}$ is the $N_r \times N_t$ channel matrix, $\mathbf{s}$ is the $N_s \times 1$ vector of transmitted symbols such that $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{1}{N_s}\mathbf{I}_{N_s}$, $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_r)$ is the Gaussian noise corrupting the received signal, and $\mathbf{I_n}$ the corresponding interference. Moreover, $\mathbf{W}$ is the $N_r \times N_s$ combining matrix. Since mm-wave channels are expected to have limited scattering [164], in this section, a geometric channel model with N_l scatterers is adopted, where each scatterer is further assumed to contribute a single propagation path between the BSs. Under this assumption, the channel $\mathbf{H}$ can be expressed as [165]

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{PL(x)}} \sum_{l=1}^{N_l} \psi_{1,l}\mathbf{a}_r(\phi_{1,l}^r)\mathbf{a}_t^H(\phi_{1,l}^t), \tag{5.5}$$

where $PL(x) = P_i h_l |x|^{-\alpha}$ denotes the path-loss between the BSs at distance x and path loss exponent α, and $\psi_{1,l}$ is the complex Rayleigh distributed gain of the l-th path such that $\mathbb{E}\left[|\psi_{1,l}|^2\right] = 1$. The angles of departure or arrival (AoDs/AoAs) are denoted by $\phi_{1,l}^t$ and $\phi_{1,l}^r$. For simplicity, it is assumed that scattering only occurs in azimuth and, thus, the BSs only implement 2D beamforming. Additionally, uniform linear arrays (ULAs) are used for the antenna response vectors, $\mathbf{a}_r(\phi_{1,l}^r)$ and $\mathbf{a}_t(\phi_{1,l}^t)$. Thus, $\mathbf{a}_t(\phi_{1,l}^t)$ can be written as

$$\mathbf{a}_t(\phi_{1,l}^t) = \frac{1}{N_t}\left[1, e^{j(2\pi/\lambda_w)d\sin(\phi_{1,l}^t)}, \ldots, e^{j(N_t-1)(2\pi/\lambda_w)d\sin(\phi_{1,l}^t)}\right], \tag{5.6}$$

where λ_w is the signal wavelength, and d is the distance between antenna elements. The array response vectors at the receiver, $\mathbf{a}_r(\phi_{1,l}^r)$, can be written in a similar fashion.

5.3.2 Performance Metrics

In this subsection, the metrics used to evaluate the performance of the base station deployment algorithm are presented.

5.3.2.1 Achievable Rate

There are several factors that support interference management in a mm-wave Massive MIMO implementation ranging from limited range, narrow beamwidths and high spatial selectivity to the use of the large number of degrees of freedom to facilitate subspace-based and interference alignment methods [166]. For mm-wave MIMO systems the efficient design of the precoders and combiners maximizes the sum-rate of the system. Thus, the achievable rate per base station can be written as follows [165]

$$R_k = \log_2\left(\det\left|\mathbf{I}_r + \frac{P_i}{N_s}\mathbf{R}^{-1}\mathbf{W}^H\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H\mathbf{W}\right|\right), \tag{5.7}$$

where the noise covariance matrix after combining, $\mathbf{R}$, is given by

$$\begin{aligned}\mathbf{R} = {} & \mathbb{E}\left\{(\mathbf{W}^H(\mathbf{n}+\mathbf{I_n}))(\mathbf{W}^H(\mathbf{n}+\mathbf{I_n}))^T\right\} - \mathbb{E}\left\{\mathbf{W}^H(\mathbf{n}+\mathbf{I_n})\right\}\mathbb{E}\left\{\mathbf{W}^H(\mathbf{n}+\mathbf{I_n})\right\}^T \\ & \mathbb{E}\left\{\mathbf{W}^H\mathbf{W}^* + \mathbf{W}^H\mathbf{I_n}\mathbf{I_n}^T\mathbf{W}^*\right\} - \mathbb{E}\left\{\mathbf{W}^H\mathbf{I_n}\right\}\mathbb{E}\left\{\mathbf{I_n}^T\mathbf{W}^*\right\}^T\end{aligned} \tag{5.8}$$

5.3.2.2 Power Consumption Model

The total energy consumption of the HetNet can be considered as the sum of the power consumed during access and backhauling operations. The total power consumed by a conventional BS during access includes the power consumed by the RF power amplifier (PA), analogue filters, as well as the power consumed by the baseband and digital signal processing. If a macro BS is considered, a fixed power consumption required for site-cooling should also be taken into

account. Meanwhile, the power consumption for backhauling includes the power consumed during the transmission to the core network and the power required to carry out the wireless cooperation. Thus, the power consumption of the network can be computed as follows

$$P_{cons} = \left(\frac{1}{\delta}P_M + P_0\right) + \sum_{k=1}^{|\mathcal{B}|}\left(\frac{1}{\delta}(f_{a,k}P_a + f_{b,k}P_b) + P_{cr} + P_{pr} + f_{w,k}P_w\right)\overline{f_{id,k}}, \quad (5.9)$$

where δ denotes the efficiency of the PA and P_a and P_b are the power consumption values for access and backhaul, respectively. P_{cr} is the power required to run the circuit components (including filters, converters and mixers), P_{pr} denotes the power for signal processing and P_0 is a constant accounting for the fixed power consumption required for control signalling and site-cooling when a macro BS is considered. Moreover, $f_{a,k}$, $f_{b,k}$, and $f_{id,k}$ are equal to one when the dynamic modes *fronthaul-backhaul* (fh-bh), *relay* or *idle* of the k-th BS are active (refer to Section 5.3.4 for further information) and zero where they are not, respectively. The notation $\bar{x}$ is used for the NOT Boolean operator. Additionally, $f_{w,k}$ is used to indicate the presence ($f_{w,k} = 1$) or absence ($f_{w,k} = 0$) of a wired connection to the core of the network and P_w is the corresponding power consumed.

5.3.2.3 Energy Efficiency

Energy efficiency metrics play a vital role in the comprehensive assessment of the energy performance of wireless communication systems providing detailed information on the energy consumption of the entire network [167]. It also provides a general framework for the quantification of energy consumption reduction achieved when dynamic operation algorithms are employed. In this section, the classical and widely-used Bit-per-Joule measurement is used to evaluate energy efficiency and defined as

$$EE = \frac{\text{Data Rate}}{\text{Power Consumption}} = \frac{\sum_{b\in\mathcal{B}_\phi} R_k}{P_{cons}} \quad \text{[bits/J]}. \quad (5.10)$$

5.3.3 Energy-Efficient BS Deployment Algorithm

Finding an optimal deployment of micro BSs which can support the peak-time wireless traffic in a cost-effective way can be an extremely complex optimization problem, due to the conflicting

interests in terms of throughput and power consumption, and computational complexity. A sub-optimal but practical approach is to choose a set $\mathcal{B}_c$ of candidate locations and select a subset of these to deploy the additional small cells such that the EE of the network is optimized.

Algorithm 2 Base Station Deployment Algorithm

1: Initialize $\mathcal{B} = \emptyset$
2: Set $\eta_{EE}(\mathcal{B}_\phi) = \eta_{EE}(\mathcal{B}_M)$
3: **while** $f_w R_{bh} < R_{fh}^{max}(1 - f_w)$ **do**
4: $\quad b = \arg\max_{b \in \mathcal{B}_c} \sum_{b \in \mathcal{B}_\phi} (C_{EE}(\mathcal{B}_\phi \cup b) - C_{EE}(\mathcal{B}_\phi))$
5: $\quad \mathcal{B} \leftarrow \mathcal{B} \cup b$
6: $\quad \mathcal{B}_c \leftarrow \mathcal{B}_c \backslash b$
7: **end while**

The heuristic greedy algorithm described in Section 5.2 is employed to select in each step the micro BS that maximizes a weighted cost function, $C_{EE}(\mathcal{B}_\phi) = \eta_{EE}(\mathcal{B}_\phi) - w_b c_b(\mathcal{B}_\phi)$. This function takes into account the EE of the system and penalizes the additional network cost, $w_b c_b(\mathcal{B}_\phi)$, for deploying and operating the BSs, with $0 < w_b \leq 1$ as the normalized cost for deploying a micro BS at an specific location and c_b a proportional cost for operating the BS. Moreover, it is assumed the existence of a macro BS located at the geographical centre of the area of interest, that provides scheduling and resource allocation for the micro BS and support to high-mobility users. The BS deployment algorithm is described in Algorithm 2.

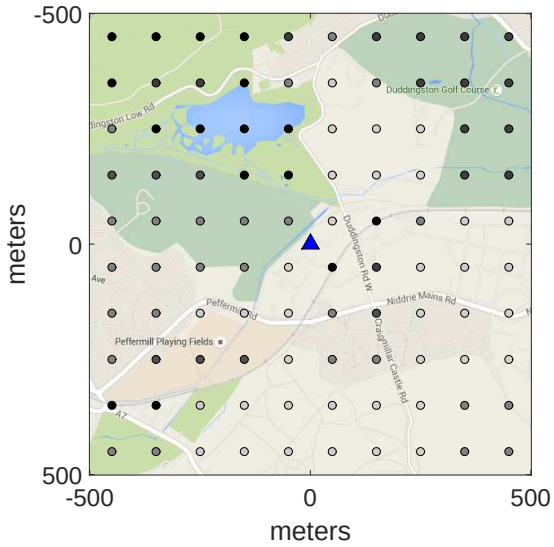


Figure 5.4: Cost associated for deploying micro BSs at each candidate location ©Google Maps, where the triangle denotes the macro BS' location and light grey and black dots show the locations with the lower and higher cost for deployment, respectively.

In Algorithm 2, $\mathcal{B}_c$ represents the set of micro BS candidates. If some locations are not available for deployment due to the local terrain, higher weights, w_b, are assigned to these locations. In this work, the area of interest is divided into an equally distributed grid, where each location has a cost associated with the suitability for BS deployment. An example of this scenario is shown in Figure 5.4, where the feasible locations for the macro and micro BS are denoted by a triangle and circles, respectively. The grey scale code indicates the cost for deploying a BS at each candidate location, where the light grey and black dots show the locations with the lowest and highest cost for deployment, respectively.

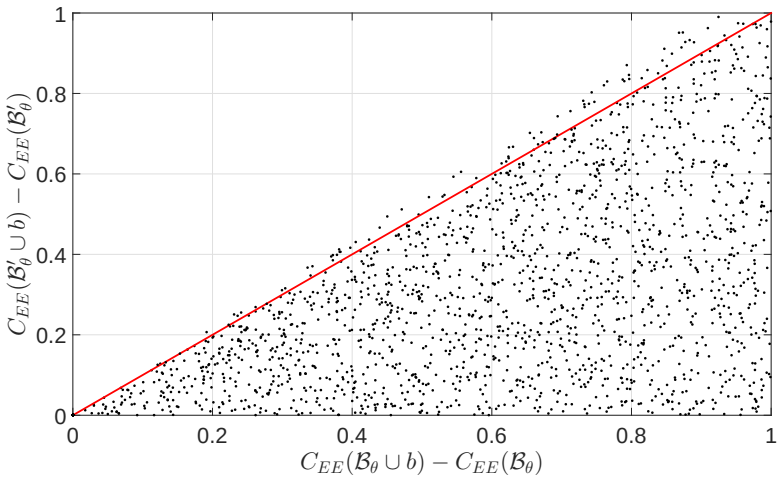


Figure 5.5: Submodularity test of the cost function $C_{EE}(B_\phi)$, where the red line, $C_{EE}(\mathcal{B}_\phi \cup b) - C_{EE}(\mathcal{B}_\phi) = C_{EE}(\mathcal{B}'_\phi \cup b) - C_{EE}(\mathcal{B}'_\phi)$, indicates the regions where the incremental value of the function increases (right) or decreases (left) as the size of the input set increases.

From Section 5.2, it is known that the greedy algorithm guarantees an optimal solution if i) $C_{EE}(\emptyset) = 0$, ii) $C_{EE}(\mathcal{B}_\phi)$ is a non-decreasing function, and iii) $C_{EE}(\mathcal{B}_\phi)$ is submodular. Even though the EE function violates condition ii), in [154] and [153], it was shown that under a fixed area and bandwidth the SE satisfies the three properties. Figure 5.5 shows that the proposed algorithm achieves submodularity, i.e., the difference in the incremental value of the function that a single element makes when added to an input set decreases as the size of the input set increases for most of the locations ($> 81\%$).

5.3.4 Traffic-Aware Dynamic BS Operation

Even when a base station experiences little to no activity, they may consume more than the 50% of their peak energy [168]. Considering that the traffic of the network is less than the 20% of

the peak-time traffic for 50% of the time, then, the BSs deployed to satisfy the requirements of the network are under-utilized most of the time [168, 169]. In order to obtain energy savings, a coordinated dynamic operation approach that allows the network to switch-off BSs and transfer their user load to different cells during periods of low utilization could be considered. Furthermore, cooperative techniques have been identified as an attractive approach to improve the EE of cellular networks [170, 171]. From the backhauling perspective, allowing some micro BSs to relay their traffic to the core of the network not only reduces the cost for operation but also ensures that shutting down micro BSs does not leave coverage holes. For this aim, it is assumed that the daily traffic profile for each type of BS follows the pattern shown in Figure 5.6, which is based on [168]. Once the network has been deployed to meet the requirements of the traffic at peak-time, a dynamic BS operation is considered to reduce the total energy consumption.

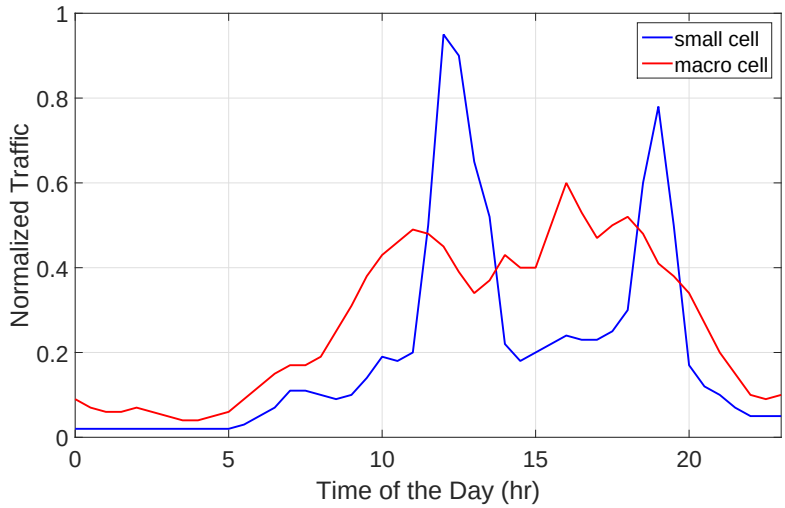


Figure 5.6: Normalized traffic load of small and macro cells regarding the time of the day.

During the dynamic or load-aware operation of the micro BSs, it is assumed that each of them can adopt one of the following modes:

- *Full operation* (full): where the BS not only performs access and backhaul operations but also has a wired connection that allows it to forward the data traffic to the core of the network;
- *Access/Fronthaul and Backhaul* (fh-bh): the BS performs wireless backhauling and access;
- *Relay*: The BS only relays the data traffic from other micro BSs;

- *Switched-off* (idle): the BS is switched-off to save energy.

Moreover, it is assumed that the number of micro BSs provided with a wired connection to the core of the network depends on the density factor, $\eta_{op} = |\mathcal{B}_{op}|/|\mathcal{B}_{\theta}|$ with $\mathcal{B}_{\theta} = \mathcal{B}_{op} \cup \mathcal{B}_{w}$, where $\mathcal{B}_{op}$ and $\mathcal{B}_{w}$ are, respectively, the sets of BSs with optical fibre and wireless backhaul connection. Then, the energy optimization problem can be written as

$$\begin{aligned} &\max_{\eta_{op},\mathcal{B}_{\theta}} P_{cons}, \\ &\text{subject to } \; f_w R_{bh} \geq R_{fh}^{t}(1 - f_w). \end{aligned} \tag{5.11}$$

Most combinatorial optimization problems, as the one described by equation (5.11), are NP-hard to solve optimally [172]. Since they cannot be solved in polynomial time unless $P = NP$, a natural approach to cope with this intractability is to use an approximation algorithm. In this section, a heuristic greedy algorithm is employed in Algorithm 3 to solve the optimization problem.

Algorithm 3 Dynamic base station operation

1: Order $\mathcal{B}_c = (\mathcal{B}_{op} \cup \mathcal{B}_w) \backslash B_M$ according the traffic load
2: Set $\mathcal{B}_d = \mathcal{B}_M$
3: Set $Q = \{full, fh-bh, relay, idle\}$ as set of the operational mode for a BS
4: **while** $f_w R_{bh} < R_{fh}^{t}(1 - f_w)$ **do**
5: $\quad b = \arg\max_{Q,\mathcal{B}_d(b)} \sum_{\mathcal{B}_d} \frac{\sum R_k(Q)}{P_{\text{cons}}(Q)}$
6: $\quad \mathcal{B}_d \leftarrow \mathcal{B}_d \cup b(Q)$
7: $\quad \mathcal{B}_c \leftarrow \mathcal{B}_c \backslash b(Q)$
8: **end while**

In Algorithm 3, two assumptions are made: additional optical fibre connections are not allowed and $R_{fh}^{t} \leq R_{fh}^{max}$. Moreover, the total power consumption $P_{cons}(Q)$ depends on the operation mode adopted by the BSs. In equation (5.9), some flags are used to indicate the energy consumption for each operational mode. Thus, if the micro BS is switched-off, $f_{id,k}$ is set to one. Otherwise, $f_{a,k}$ and $f_{b,h}$ are activated according to the operation mode. Notice that in (5.9), the power consumption for signal processing and running the circuit are always considered when the BS remains active. Since the aim of Algorithm 3 is to maximize the $EE = \frac{\sum R_k(Q)}{P_{\text{cons}}(Q)}$ performance of the system, Section 5.5 presents the numerical results obtained by using Algorithms 2 and 3 for base station deployment and dynamic operation, respectively.

In the next section, the associated trade-offs of an energy-efficient self-backhauling heteroge-

neous network with inhomogeneous distribution of the users are investigated.

5.4 Sleeping Strategies for Inhomogeneous Cellular Network

A well-researched method to model HetNets is the assumption of independent homogeneous Poisson point processes to model the locations of both users and base stations. Despite the analytical tractability this approach lacks accuracy to model future 5G wireless networks, since base stations are usually deployed in locations where the users are more likely to appear, so the UEs will be clustered towards the BS, rather than being uniformly distributed. In this section, the associated trade-offs for the downlink of an energy-efficient self-backhauling heterogeneous network with inhomogeneous user distribution are analysed through the employment of sleeping strategies. Three different policies for switching-off base stations are considered: random, load-based and greedy algorithms. Moreover, the probability of coverage for the random and load-based sleeping policies and study of the performance of the heuristic greedy algorithm are presented.

5.4.1 System Model

Consider a heterogeneous cellular network where a set of micro BSs, $\mathcal{B}$, are assumed to transmit at the same power P. The locations of the BSs are modelled by an independent homogeneous PPP, $\Phi_1^{(BS)}$, of density $\lambda_1 > 0$. A fraction ω of the BSs, called anchored BS (A-BS) henceforth, have a wired connection to the core of the network while the rest backhaul wirelessly to A-BSs; this results in single-hop and two-hop links to the users associated with the A-BSs and BSs, respectively (see Figure 5.7) [173]. Thus, a Bernoulli random variable with probability ω is used to assign a wired backhaul (or not) to each BS. Unlike prior works which were mainly focused on the backhaul performance analysis when the UEs are uniformly distributed, in this section, the analysis is focused on scenarios with a non-homogeneous user distribution.

As mentioned in [141], while the homogeneous PPP assumption for the UE distribution is not practical, highly clustered distributions with no UEs in-between are also unlikely in practice. Thus, a tunable model that covers a wide range of scenarios from a highly homogeneous PPP to a highly heterogeneous clustered case is considered in this section. As a result, users can be divided into two main categories: (a) users that are correlated with the BSs or points of interest and (b) users that are not correlated with the points of interest, i.e., the distribution of these

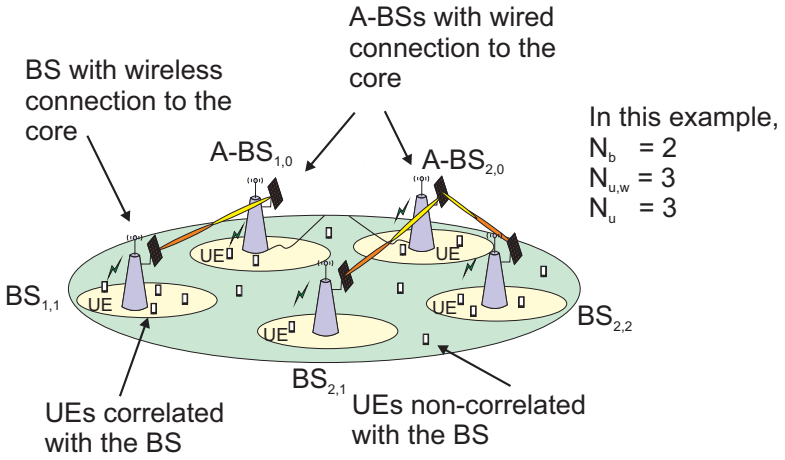


Figure 5.7: Diagram of the HetHetNet wireless network considered.

users is independent from the location of the BSs.

In particular, the distribution of UEs can be considered as the superposition of two independent distributions. The first one is a homogeneous PPP with density $p\lambda^{(UE)}$ and the second one is a cluster process with total density of $(1-p)\lambda^{(UE)}$ [141]. Thus, in a network with $p = 0$, all the users locations are correlated with the points of interest, and the resulting network is a clustered point process. On the contrary, a PPP generated with $p = 1$ models a uniform user distribution in space. Any other value of $0 < p < 1$ results in the superposition of a homogeneous PPP and a Thomas cluster process (see Figure 5.8).

When the users are correlated ($p = 0$) with a given BS located at $\mathbf{x}_i \in \mathbb{R}^2$, then, the users are assumed to be symmetrically, independently, and identically distributed (i.i.d.) around $\mathbf{x}_i \in \Phi_1^{(BS)}$. Thus, a *Poisson cluster process* denoted by Φ_1^u is formed by the union of all such locations of users. In other words, the points of $\Phi_1^{(BS)}$ serve as cluster centres for Φ_1^u. Moreover, it is assumed that the user location $\mathbf{z} \in \mathbb{R}^2$ with respect to its cluster centre, $\mathbf{x}_i$, follows a probability density function (PDF) $f(\mathbf{z})$. Following [147], the PP Φ_1^u is modelled as a Thomas cluster process, where the cluster members (or UEs) are scattered according to a symmetric normal distribution with variance σ^2 around the BS. Thus, [15]

$$f(\mathbf{z}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\parallel \mathbf{z} \parallel^2}{2\sigma^2}\right), \qquad \mathbf{z} \in \mathbb{R}^2. \tag{5.12}$$

Without loss of generality and according to the independence property of the PPPs, which suggests that conditioning the process on a randomly chosen point of the cluster does not change

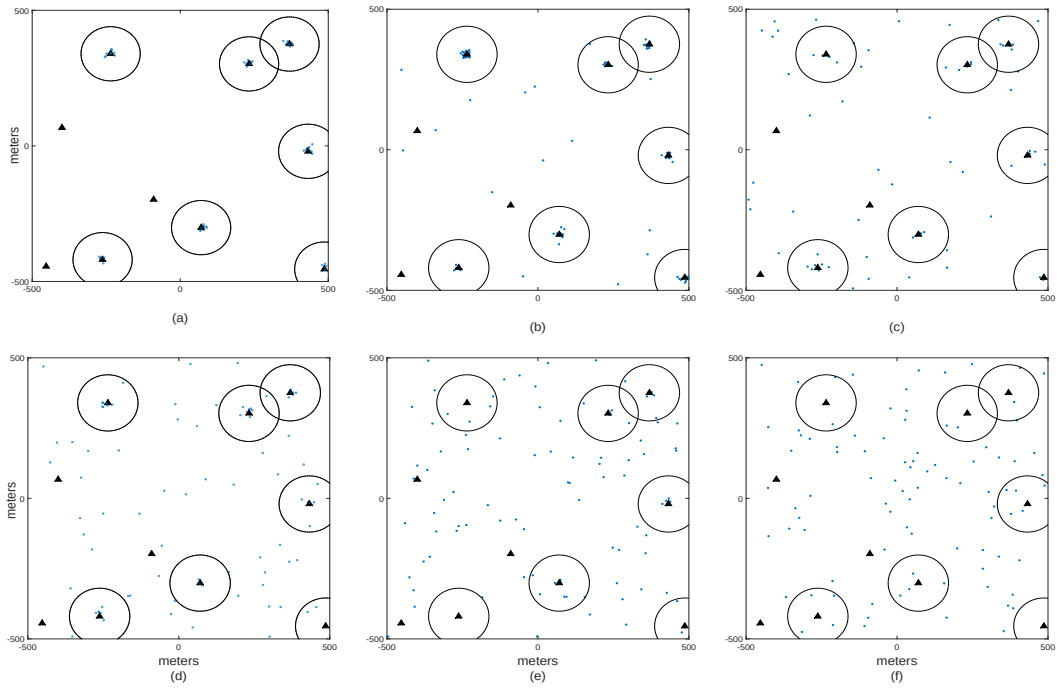


Figure 5.8: Various UE distributions in a 1000m $\times$1000m network with different clustering properties: (a) $p = 0$, (b) $p = 0.20$, (c) $p = 0.40$, (d) $p = 0.60$, (e) $p = 0.80$ and (f) $p = 1.00$. Starting from a pure Thomas cluster process until a homogeneous PPP. The number of BSs is 10 and the cluster variance is 100m. Black triangles denote the BSs and small blue dots denote the UEs.

its distribution, the downlink analysis is performed at a typical user of Φ_1^u [15]. Since the PPPs are stationary processes, then the origin can be transformed to the location of the typical user. Thus, given that the typical user is at the origin, the location of the cluster centre is $\mathbf{y} \in \Phi_1^{(BS)}$. Slivnyak's theorem guarantees that if this point is excluded from $\Phi_1^{(BS)}$, i.e., $\Phi_1^{(BS)} \setminus \{\mathbf{y}\}$, the remaining process has the same distribution as $\Phi_1^{(BS)}$ [15, 147].

For simplicity and following the notation suggested in [147], an additional tier, $\Phi_0^{(BS)}$, is formed. This tier consists of only of the cluster centre, i.e., $\Phi_0^{(BS)} = \{\mathbf{y}\}$. As a result, the set of indices of this 2-tier HetNet can be defined as $\mathcal{K} = \mathcal{K}_0 = \{0\} \cup \mathcal{K}_1 = \{1\}$. Thus, the user can either connect to its own cluster centre, i.e., the BS in $\Phi_0^{(BS)}$, or, to some other BS $\in \Phi_1^{(BS)}$. Assuming that the typical user is served by the BS at $\mathbf{x}_j \in \Phi_k^{(BS)}$ with $k \in \mathcal{K}$, the received power can be expressed as

$$P(\mathbf{x}_j) = Ph_j \parallel \mathbf{x}_j \parallel^{-\alpha}, \tag{5.13}$$

where α is the path loss exponent and P is the transmit power. In order to incorporate the ran-

dom effects of the channel a multiplicative random value h_j is considered. Under the Rayleigh fading assumption, the channel coefficients h_j are i.i.d. exponential random variables with unit mean. Throughout this chapter, the notation $h_j \sim \exp(1)$ describes an exponential random variable with unit mean.

Average power-based cell selection in which a user (or BS) connects to the BS (or A-BS) that provides maximum received power is assumed. Considering that the average power from a BS at $\mathbf{x}_j \in \Phi_k^{(BS)}$ is $P \parallel \mathbf{x}_j \parallel^{-\alpha}$, then the set of serving BSs is the set of closest BS from each tier [174].

5.4.2 Performance Metrics

In this subsection, the metrics used to evaluate the performance of the sleeping strategies are presented.

5.4.2.1 Access and Backhaul Load

From the above discussion, access and backhaul links are assumed to share the available radio resources (or bandwidth) and hence the throughput of the network depends on the user load at the BSs/A-BSs. Let N_b, $N_{u,w}$, and N_u denote the number of BSs with wireless backhaul links to the tagged A-BS, the number of users served by the tagged A-BS, and the number of users associated with the tagged BS, respectively.

Since an A-BS serves both users and BSs, the radio resources allocated to the associated BSs are assumed to be proportional to their user load. Every user is labelled as BS-independent with probability p or BS-correlated with probability $(1-p)$. The average number of independent users per BS is $\kappa_1 = p\lambda^{(UE)}/(\lambda_1+\lambda_0)$, meanwhile, the average number of correlated users per BS is $\kappa_0 = (1-p)\lambda^{(UE)}/(\lambda_1+\lambda_0)$. Thus, the fraction of radio resources available for all the associated BSs at an A-BS are

$$\eta_b = \frac{(\kappa_0+\kappa_1)N_b}{(\kappa_0+\kappa_1)N_b + N_{u,w}}, \tag{5.14}$$

and the fraction for the access link with the associated users are

$$\eta_a = 1 - \eta_b = \frac{N_{u,w}}{(\kappa_0 + \kappa_1)N_b + N_{u,w}}. \tag{5.15}$$

Generally speaking, the number of UEs clustered towards each BS may be different due to strategic base station locations and transmit power. Nevertheless, in this work, it is assumed that the number of dependent users per BS is uniform across the network.

Moreover, the fraction of bandwidth available for the BSs served by an A-BS are assumed to be shared equally among the BSs and hence the fraction of radio resources available to the tagged BS from the tagged A-BS are η_b/N_b, which is equivalent to the resource fraction used for backhaul by the corresponding BS. The access and backhaul capacity at each BS is assumed to be shared equally among the associated users [173] .

5.4.2.2 SINR Model

Considering the cooperative system model described in Section 5.4.1, the value of SINR at the typical receiver when it is served by the BS or A-BS (see Figure 5.9 located at a distance $\| \mathbf{x} \|$ can be described as [173],

$$\mathit{SINR}_a = \frac{PG_BG_Uh_a \| \mathbf{x}_a \|^{-\alpha}}{I_a + \sigma_N^2}, \qquad \mathit{SINR}_b = \frac{PG_B^2h_b \| \mathbf{x}_b \|^{-\alpha}}{I_b + \sigma_N^2} \tag{5.16}$$

where $\sigma_N^2 = N_0B$ is the thermal noise power, h_a and $h_b \sim \exp(1)$, G_B and G_U denote the antenna gain of the BS and user, respectively. Recall, when a interference limited system is considered $\sigma_N^2 = 0$ and (5.16) reduces to the SIR parameter. Furthermore, I_a and I_b are the corresponding interference for access and backhaul links, respectively. These terms are dependent on the service distance and details on how to calculate them are provided in Section 5.4.3.

5.4.2.3 Throughput and Data Rate

A user cannot be served by a BS unless the received SINR reaches a given threshold, i.e., $\mathbb{P}_C(\mathit{SINR} > \tau)$, where τ is a prescribed quality-of-service (QoS) SINR value. Thus, the

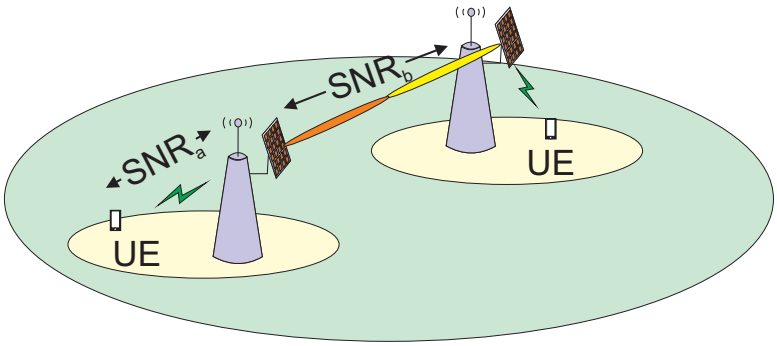


Figure 5.9: SINR model of the HetHetNet described in Section 5.4.1.

throughput attained at a given BS - UE link is given by

$$T = \mathbb{P}_C(\mathit{SINR} > \tau)\log_2(1+\tau). \tag{5.17}$$

Considering the resource allocation model (eqs. (5.14) and (5.15)), the SINR equation in (5.16) and the achievable rate for the DF protocol in equation (2.22), the rate for a typical user is given by [173]

$$R_k = W\log_2(1+\tau) = \begin{cases} \frac{W}{N_{u,w}+(\kappa_0+\kappa_1)N_b}\log_2(1+\mathit{SINR}_a), \text{if associated with a wired A-BS,} \\ \frac{W}{N_u}\min\left\{\left(1-\frac{(\kappa_0+\kappa_1)}{(\kappa_0+\kappa_1)N_b+N_{u,w}}\right)\log_2(1+\mathit{SINR}_a), \right. \\ \qquad \left. \frac{(\kappa_0+\kappa_1)}{(\kappa_0+\kappa_1)N_b+N_{u,w}}\log_2(1+\mathit{SINR}_b)\right\}, \text{otherwise} \end{cases} \tag{5.18}$$

where W is the bandwidth of the system. Notice, if the typical user is served by a BS, due to the two-hop cooperative link, then the rate is limited by the minimum rate between the access and backhaul links, as in the lower part of (5.18).

5.4.2.4 Power Consumption Model

As mentioned in Section 5.3.4, the traffic of the network is less than the 20% of the peak-time traffic for 50% of the time. Thus, the network becomes under-utilized most of the time. Rewriting the power consumption model presented in Section 5.3.2.2, the power consumed by the network can be calculated as follows

$$P_{cons} = \sum_{k=1}^{|\mathcal{B}|} \left(\frac{1}{\delta}(P + f_{b,k}P) + P_{cr} + P_{pr} \right), \tag{5.19}$$

where δ denotes the efficiency of the PA, P_{cr} is the power required to run the circuit components (including filters, converters and mixers), and P_{pr} denotes the power for signal processing. The variable $f_{b,k}$ indicates if the k-th micro BS performs a wireless ($f_{b,k} = 1$) or wired ($f_{b,k} = 0$) backhaul operation.

5.4.2.5 Energy Efficiency

Theoretically, in order to compare the performance of two systems, it is enough to set a target throughput and measure the power drawn by each of the systems. Nevertheless, discrepancies on the systems' capabilities and cost to run the tests make normalizing the energy consumption by the effective throughput more suitable for practical implementation. Thus, in this section, the reciprocal of Bit-per-Joule measurement, namely energy consumption per delivered information bit, referred to as the energy consumption ratio (ECR) [169] is used and defined as

$$ECR = \frac{\text{Power Consumption}}{\text{Data Rate}} \text{ [J/bit]}. \tag{5.20}$$

5.4.3 Probability of Coverage

In this section, the effect of the inhomogeneous user distribution in the performance of the network is investigated. The analysis is mainly focused on the distance distribution between a typical user and its tagged A-BS or BS and its impact on the probability of coverage. Recall that a user cannot be served by a BS unless the received SINR reaches a desired threshold. The probability of coverage can be defined as $\mathbb{P}_C(\textit{SINR} > \tau)$ [77], which can be seen as the probability that a randomly chosen user can achieve a target SINR τ [72].

To begin, let R_1 be a random variable denoting the distance from a typical user at the origin to the nearest point of the Poisson point process $\Phi_1^{(BS)}$. Since the typical user is served by the closest base station, no other base station can be closer than R_1. Thus, the PDF of R_1 can be derived using the fact that the null probability of a spatial Poisson process within an area A of

interest is $\exp(-\lambda_1 A)$ [71]

$$\begin{aligned}\mathbb{P}(r > R_1) &= \mathbb{P}(\text{No BS closer than } R_1) \\ &= \exp(-\lambda_1 \pi R_1^2). \end{aligned} \tag{5.21}$$

From the above discussion, it is known that the typical user is served by a BS located at a distance R_1. Since, the probability that two BSs lie on this boundary is zero [73], there exists no other BS within the disc $b(0, R_1)$ of radius R_1. The notation $b(c, x)$ is used to describe a disc centred at c of radius x. Accordingly, the PDF and cumulative density function (CDF) of R_1 are given by [15]

$$f_{R_1}(r) = 2\pi r \lambda_1 \exp(-\pi \lambda_1 r^2), \qquad r \geq 0, \tag{5.22a}$$

$$F_{R_1}(r) = 1 - \exp(-\pi \lambda_1 r^2), \qquad r \geq 0. \tag{5.22b}$$

Following the notation introduced in Section 5.4.1, the distribution of R_0 needs to be treated differently as $\Phi_0^{(BS)}$ contains only a single point with a predefined distance distribution. Assuming that Φ_0^u is a *Thomas cluster process*, R_0 is Rayleigh distributed with PDF and CDF given by [147, 175, 176]

$$f_{R_0}(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \qquad r \geq 0, \tag{5.23a}$$

$$F_{R_0}(r) = 1 - \exp\left(-\frac{r^2}{2\sigma^2}\right), \qquad r \geq 0. \tag{5.23b}$$

Recalling that the superposition of two independent Poisson processes results in a new Poisson process where its density equals the sum of the individual densities, then the PDF and CDF of the superposition of an homogeneous and clustered distributions can be computed as

$$f_p(r) = pf_{R_1} + (1-p)f_{R_0} \tag{5.24a}$$
$$= 2\pi r\lambda_1 p\exp(-\pi\lambda_1 r^2) + \frac{1}{\sigma^2}\left(r(1-p)\exp\left(-\frac{r^2}{2\sigma^2}\right)\right), \qquad r \geq 0$$
$$F_p(r) = pF_{R_1} + (1-p)F_{R_0} \tag{5.24b}$$
$$= 1 + (p-1)\exp\left(-\frac{r^2}{2\sigma^2}\right) - p\exp(-\lambda_1\pi r^2), \qquad r \geq 0,$$

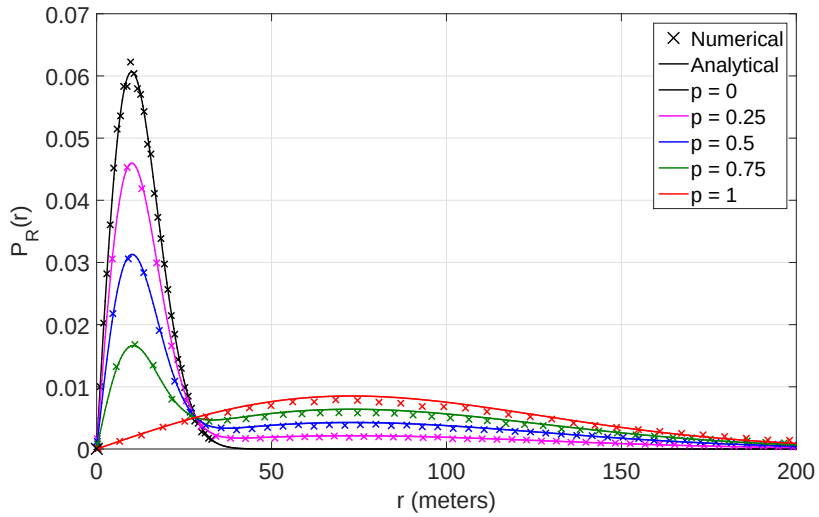


Figure 5.10: Probability density function of the serving distance between the typical user and its tagged BS for NumBS = 30, NumUE = 1000, Area = 1000m $\times$1000m and the superposition probability $p = 0, 0.25, 0.5, 0.75$ and 1.

Figure 5.10 shows the PDF of the serving distance between a typical user and its tagged BS. It is clear from the plot that analytical and numerical results match and that the superposition probability p plays a vital role in determining the performance of the HetHetNet. Before the probability of coverage of the HetHetNet is formally introduced, the effect of the inhomogeneous user distribution is studied by calculating the ratio of the distances from a typical user to its tagged A-BS or BS (R_S) and to the second closest BS, i.e., its dominant interferer (R_I). Following the approach in [151] and assuming a interference-limited system, the SIR, γ, in dB can be calculated as

$$\gamma = P_S + G - I_{agg}, \tag{5.25}$$

where $G = G_U G_B$ is the antenna gain, the transmit power associated with the tagged A-BS or

BS is given by

$$P_S = P + G - PL_S + h_S, \tag{5.26}$$

and the aggregate interference is

$$I_{agg} = \sum_{j \in \Phi_k^{(BS)} \setminus S} (P + G - PL_j + h_j), \tag{5.27}$$

with h_S and h_j following a normal Gaussian distribution $\mathcal{N}(\mu, \sigma_N^2)$. Thus, the SIR can be upper-bounded considering only the strongest interferer as follows

$$\gamma = P_S - I_{agg} \leq \hat{\gamma} = P_S - P_I \tag{5.28a}$$

$$= PL_S - PL_I + h, \tag{5.28b}$$

where $PL_i[dB] = 10\alpha \log_{10}(R_i)$ with $i \in \{S, I\}$ is the path loss and $h = h_S + h_I \sim \mathcal{N}(2\mu, 2\sigma_N^2)$. Furthermore,

$$\hat{\gamma} = -\frac{10}{\log(10)} \cdot \alpha \cdot \log\left(\frac{R_S}{R_I}\right) + h \tag{5.29a}$$

$$\hat{\gamma} = -\frac{10}{\log(10)} \cdot \alpha \cdot \log(R_{S,I}) + h, \tag{5.29b}$$

where $0 \leq R_S \leq R_I$, and as result $R_{S,I} \in [0, 1]$ has the following CDF and PDF

$$F_r(R_{S,I}) = (1-p)\frac{R_{S,I}^2}{R_{S,I}^2 - 2\pi\lambda_1(R_{S,I}^2 - 1)s^2} + pR_{S,I}^2, \qquad 0 \leq R_{S,I} \leq 1 \tag{5.30a}$$

$$f_r(R_{S,I}) = 2pR_{S,I} + (1-p)\left(-\frac{R_{S,I}^2(2R_{S,I} - 4\pi\lambda_1 R_{S,I}\sigma^2)}{(R_{S,I}^2 - 2\pi\lambda_1(R_{S,I}^2 - 1)\sigma^2)^2} + \frac{2R_{S,I}}{R_{S,I}^2 - 2\pi\lambda_1(R_{S,I}^2 - 1)\sigma^2}\right), \qquad 0 \leq R_{S,I} \leq 1. \tag{5.30b}$$

Further details of the derivation of (5.30) are presented in Appendix A.

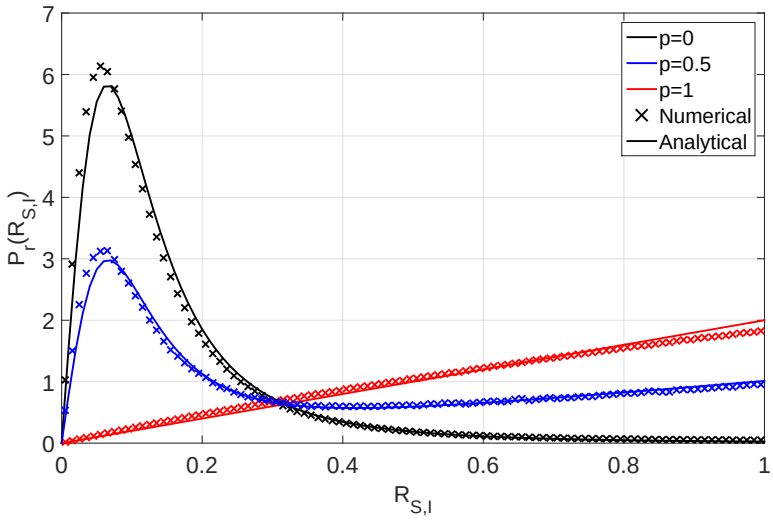


Figure 5.11: Probability density function of the ratio $R_{S,I} = R_S/R_I$.

Figure 5.11 shows the analytical results closely match the numerical results for the PDF of the ratio $R_{S,I} = R_S/R_I$. It can be seen that the average of $R_{S,I}$ is lower for a highly clustered system compared with the average obtained with a homogeneous distribution. From this observation it can be pointed out that the distance between the first and second nearest BS increases as $p \to 0$. Moreover, there is a crossing point at $R_{S,I} = 0.31$ that is essentially defined by the cluster variance and the density of the homogeneous distribution. Furthermore, in terms of interference the smaller the ratio $R_{S,I}$, the lower the interference a typical user experiences. Finally, this crossing point shows where the cluster or homogeneous distribution becomes more dominant in the system's performance.

For the sake of clarity and using the inverse transform theorem, the PDF of the logarithm of $R_{S,I}$, i.e., $Y = \log(R_{S,I})$ can be obtained as

$$f_Y(y) = e^y \left[2pe^y + (1-p) \left(-\frac{e^{2y}(2e^y - 4\pi\lambda_1 e^y \sigma^2)}{(e^{2y} - 2\pi\lambda_1(e^{2y}-1)\sigma^2)^2} + \frac{2e^y}{e^{2y} - 2\pi\lambda_1(e^{2y}-1)\sigma^2} \right) \right], \quad y \leq 0. \quad (5.31)$$

For further details of the derivation of (5.31), see Appendix A.

The analytical and numerical results of (5.31) are shown in Figure 5.12. In this analysis, the

discussion will be focused on the case where $p = 0.5$, since the results for the inhomogeneous case in Section 5.5.2 are closely related to performance exhibited in this plot. As can be seen, unlike the homogeneous case ($p = 1$) where the PDF of $R_{S,I}$ follows an exponential growth or the highly clustered network ($p = 0$) that follows a log-normal distribution, when the superposition of the two user distributions occurs, the PDF of $R_{S,I}$ has a peak around -2 dB. This can be explained considering the crossing point in Figure 5.11. When the probability of $R_{S,I} \to 0$ the stronger interferer is located far away from the serving BS and thus it is highly likely that the typical user and the tagged BS are correlated. Nevertheless, when $R_{S,I}$ approaches the crossing point at 0.31, the typical user can be served either by its cluster centre or another BS. As can be seen from Figure 5.12, on the left hand side of the crossing point the inhomogeneous distribution of the network becomes dominant, but the opposite applies if instead the right hand side is considered. The oscillatory behaviour at -2 dB can be attributed to the different weights that each of the distributions impose when the serving distances are varied. As will be seen in Section 5.5.2, the load-based algorithm, where the serving distance is not a crucial decision parameter, is going to be penalized in terms of overall throughput. In general, the performance of the HetHetNet is heavily dependent of the cluster variance, the density of BSs in the area of interest and the inhomogeneity of the user distribution.

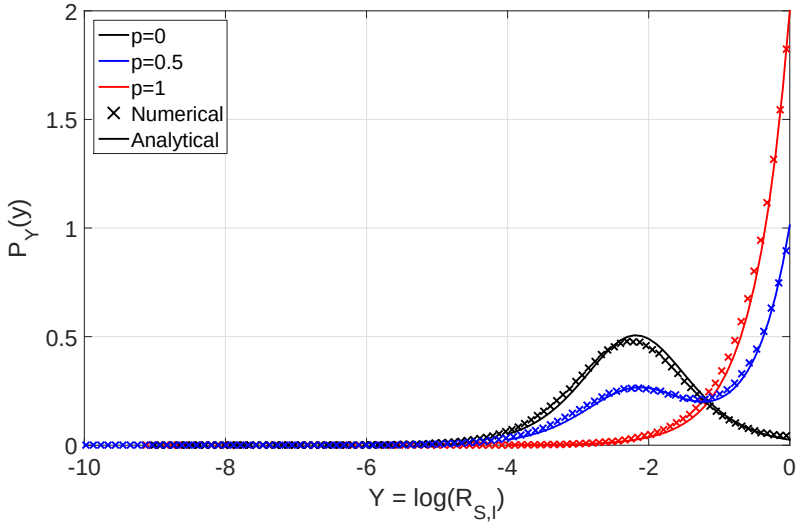


Figure 5.12: PDF of the $\log(R_{S,I}) = \log(R_S/R_I)$.

From the discussion above, it can be deduced that if a typical user associates with a BS of location $\mathbf{x}_j \in \Phi_1^{(BS)}$, then the total interference experienced by the typical user is generated from two independent sets: (1) the centre of its cluster, $\mathcal{I}_0 = Ph_0 \parallel \mathbf{x}_0 \parallel^{-\alpha}$ and (2) the cumulative interference from all the other BSs excluding the tagged BS, in other words, $\mathcal{I}_1 =$

$\sum_{\mathbf{x}_k \in \Phi_1^{(BS)} \setminus \mathbf{x}_j} P_k h_k \parallel \mathbf{x}_k \parallel^{-\alpha}$. Moreover, if the typical user is served by its cluster centre, the interference $\mathcal{I}_H$ is generated by the remaining BSs, i.e., $\sum_{\mathbf{x}_k \in \Phi_1^{(BS)} \setminus b(0,x_0)} P_k h_k \parallel \mathbf{x}_k \parallel^{-\alpha}$. Thus, I_a and I_b in (5.16) can be calculated based on the terms $\mathcal{I}_0$, $\mathcal{I}_1$ and $\mathcal{I}_H$.

Considering the fact that under the Rayleigh fading assumption all the sources of interference are independent of each other and using the Campbell-Mecke theorem, the probability of coverage can be expressed as a product of their Laplace transforms [177, 178]. Thus, the overall coverage probability can be defined in terms of the per-tier coverage probability as follows

$$\mathbb{P}_C = p \cdot \mathbb{P}_{C_H} + (1-p) \cdot \mathbb{P}_{C_T}, \tag{5.32}$$

where the values $\mathbb{P}_{C_H}$ and $\mathbb{P}_{C_T}$ account for the probability of coverage of the homogeneous and the Thomas cluster processes cases, respectively. From [147], it is known that the

$$\mathbb{P}_{C_H} = \int_{r>0} \prod_{k \in \mathcal{K}} \mathcal{L}_{I_H}(\tau r^{\alpha} P^{-1}) f_R(r) dr, \tag{5.33a}$$

where $\mathcal{L}_{I_H}(s) = \mathbb{E}[-sI_H]$ denotes the Laplace transform of the interference terms and $f_R(r)$ the PDF of the distance distribution. Given that the BS' locations follow a homogeneous PPP distribution, the Laplace transform of $\mathcal{L}_{I_H}$ evaluated at $\tau r^{\alpha} P^{-1}$ is

$$\mathcal{L}_{I_H}(\tau r^{\alpha} P^{-1}) = \exp(-\pi(\lambda_0 + \lambda_1)\rho(\tau, \alpha) r^2), \tag{5.34}$$

where $\rho(\tau, \alpha) = \frac{2\tau}{\alpha - 2}\, {}_2\mathcal{F}_1[1, 1 - 2/\alpha; 2 - 2/\alpha, -\tau]$ is the Gaussian Hypergeometric function. On the other hand, considering the inhomogeneity of the network, the probability of coverage corresponding to the Thomas cluster process, $\mathbb{P}_{C_T}$ can be written as

$$\mathbb{P}_{C_T} = \mathbb{P}_{C0} + \mathbb{P}_{C1}, \tag{5.35}$$

where $\mathbb{P}_{C0}$ and $\mathbb{P}_{C1}$ are the probabilities of coverage associated to the BS $\in \Phi_0^{(BS)}$ and $\Phi_1^{(BS)}$, respectively. From [147, 152], the Laplace transforms from the different components of the clustered Poisson process are given by

$$\mathcal{L}_{I_0}(\tau r^{\alpha} P^{-1}) = \exp(-\pi(\lambda_0 + \lambda_1)\rho(\tau, \alpha) r^2), \tag{5.36}$$

$$\mathcal{L}_{I_1}(s) = \int_{x_i}^{\infty} \frac{1}{1+sPr^{-\alpha}} \frac{f_{R_0}(r)}{1-F_{R_0}(x_i)} \mathrm{d}r, \tag{5.37}$$

where (5.36) deals with the interference from all the BSs, except the cluster centre. Recall that since the 0^{th} tier is created where the typical user is located, the exclusion zone will be the ball $b(0, x_0)$. In the case of (5.37), a virtual exclusion zone is formed around the typical user, since the cluster centre acting as a interferer lies outside this ball. As a result, the distance distribution $f_{R_0}(r)$ will be conditioned on $R_0 > x_0 = x_i$. The per-tier probability of coverage $\mathbb{P}_{C0}$ and $\mathbb{P}_{C1}$ can be described as [147]

$$\mathbb{P}_{C0} = \frac{\lambda_0}{\rho(\tau,\alpha)\lambda_1 + \lambda_0}, \tag{5.38}$$

$$\begin{aligned} \mathbb{P}_{C1} = 2\pi\lambda_1 \int_0^{\infty} \int_{x_i}^{\infty} \left[\frac{1}{\sigma^2} \frac{r}{1+\tau(r/x_i)^{-\alpha}} \exp\left(\frac{-r^2+x_i^2}{2\sigma^2} \right) dr \right] \\ \times x \cdot \exp\left(-\pi x^2 (1+\rho(\tau,\alpha))\lambda_1 + \lambda_0) \right) dx. \end{aligned} \tag{5.39}$$

Once the inhomogeneity of the network and its impact on the probability of coverage have been accounted for, the next section aims to improve the system's performance by using sleeping strategies to reduce the power consumption during low activity periods.

5.4.4 Sleeping Strategies

Since full operation of the micro BSs is not required at all times and all locations due to the inhomogeneous distribution of traffic over time and space, in the next subsections, three policies to optimize the power consumption and throughput of the network are presented. The effect of dynamically switching-off micro BSs based on a random policy and two deterministic policies (load-based and greedy algorithms) is investigated. The system's performance is quantified by using two metrics: probability of coverage and ECR. Previous works, studying the coverage in homogeneous scenarios shown that the probability of coverage is independent of the transmit power in the interference-limited regime, i.e., $\sigma^2 \longrightarrow 0$ [155]. This effect is a result of increasing the density of the network by adding more BSs and therefore not only improving the received signal but also increasing the interference to the tagged user. As it is shown in the

following subsections, this effect can be more evident and severe as the network becomes more inhomogeneous.

5.4.4.1 Random Sleeping

This sleeping strategy can be modelled as a Bernoulli trial [155] such that each micro BS operates with probability q and is turned off with probability $(1-q)$, independent of all the other base stations' load or the throughput of the network. Therefore, after applying random sleeping over the micro BSs within the area of interest, the average total power consumption of the network can be calculated as

$$P_{RS} = q(\lambda_1 \cdot P_{cons}) + (1-q)(\lambda_1 \cdot P_0), \tag{5.40}$$

where P_{cons} and P_0 are the consumed power of each micro BS when it is active and idle, respectively. In inhomogeneous HetHetNets with random sleeping, the coverage probability is given by

$$\begin{aligned}\mathbb{P}_{C_{RS}} = (1-p)\Bigg\{ & \int_{x>0} 2\pi q\lambda_0 r \exp(-\pi q\lambda_0(\rho(\tau,\alpha)+1)r^2)\exp(\tau r^\alpha \sigma_N^2 P^{-1})dr \\ & + 2\pi q\lambda_1 \int_0^\infty \int_{x_i}^\infty \left[\frac{1}{\sigma^2}\frac{r}{1+\tau(r/x_i)^{-\alpha}}\exp\left(\frac{-r^2+x_i^2}{2\sigma^2}\right)dr\right] \\ & \times\, x\cdot\exp\left(-\pi x^2(1+\rho(\tau,\alpha)q\lambda_1+\lambda_0)\right)dx \Bigg\} \\ & + p\left\{\int_0^\infty 2\pi q\lambda_1 r\exp(-\pi q\lambda_1(\rho(\tau,\alpha)+1)r^2)\exp(\tau r^\alpha\sigma_N^2 P^{-1})dr\right\}.\end{aligned} \tag{5.41}$$

When the background noise is not considered, $\sigma_N^2 = 0$, it reduces to

$$\mathbb{P}_{C_{RS}} = (1-p)\left(q\left(\mathbb{P}_{C0}+\mathbb{P}_{C1}\right)+\frac{1-q}{1+\rho(\tau,\alpha)}\right)+p\frac{1}{1+\rho(\tau,\alpha)}, \tag{5.42}$$

where $\mathbb{P}_{C0}$ and $\mathbb{P}_{C1}$ are the probabilities of coverage described in equations (5.38) and (5.39), respectively. Compared with the results in [155], where a homogeneous system is considered, with an inhomogeneous distribution every time a centre of a cluster is switched-off, its load is shifted to surrounding BSs, changing the percentage of BS-correlated users. Moreover, if p is set to one for the scenario under consideration, equation (5.42) reduces to the results presented

in [155]. Figure 5.13 shows the probability of coverage achieved by the random sleeping algorithm when an arbitrary QoS τ is considered. As can be seen, when $\tau < -20$ dB all the users achieved the desired QoS. Nevertheless, as the target τ increases, the probability that the SIR parameter achieves this value reduces and thus the probability of coverage tends to zero for $\tau \geq 50$ dB. Finally, Figure 5.13 clearly indicates that the analytical curves match the numerical results obtained through simulation.

Proof: See Appendix B.

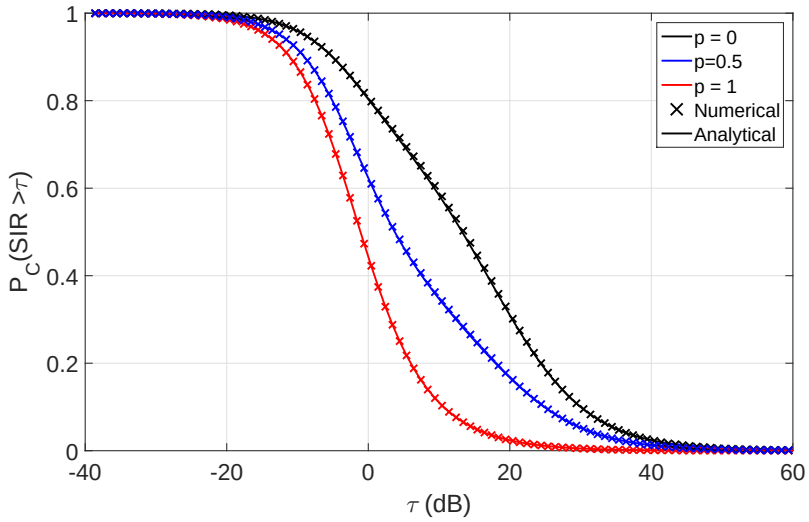


Figure 5.13: Probability of coverage when a random sleeping is employed for $q = 0.7$, $\sigma_N^2 = 0$ and $p = 0, 0.5$ and 1.

5.4.4.2 Strategic Sleeping 1: Balancing the Load - A Voronoi Based Scheme

Instead of randomly switching micro BSs off, they can be switched-off based on their activity levels. This sleeping strategy can be seen as a load-aware policy that incorporates a traffic profile in the optimization problem. Specifically, this strategic sleeping is modelled as a function $s : x \longmapsto [0, 1]$ which implies that if the activity level of the coverage area associated with the BS has activity level $x \in [0, 1]$, then it operates with probability $s(x)$ and sleeps with probability $(1 - s(x))$, independently [155]. In other words, a user located in a Poisson-Voronoi cell of activity level a is active with the same probability a, and thus a cell with a higher level of activity will exhibit a higher probability of remaining active. As a result, the average power consumption of the HetHetNet when a load-aware sleeping strategy is employed is given by

$$P_{LB} = \mathbb{E}[a](\lambda_1 \cdot P_{cons}) + (1 - \mathbb{E}[a])(\lambda_1 \cdot P_0), \tag{5.43}$$

where $\mathbb{E}[a] = \int a(x) f_A(x) dx$ and $f_A(x)$ is the pdf of the variable A, which denotes the activity within the Poisson-Voronoi cell. Moreover, if the probability of coverage is conditioned on the activity of a typical cell, then the probability of coverage for the inhomogeneous case presented can be described as follows

$$\mathbb{P}_{C_{LB}} = \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x \mathbb{P}(\textit{SINR} > \tau | x) f_A(x) dx \tag{5.44a}$$

$$= \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x s(x) \mathbb{P}(\textit{SINR} > \tau | N = 1) f_A(x) dx + \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x(1 - s(x)) \mathbb{P}(\textit{SINR} > \tau | N > 1) f_A(x) dx, \tag{5.44b}$$

where $\mathbb{P}(\textit{SINR} > \tau | N = i)$ is the probability of coverage for the typical user when it is served by the i-th BS. While the probability $\mathbb{P}(\textit{SINR} > \tau | N = 1)$ was defined by equation (5.32), $\mathbb{P}(\textit{SINR} > \tau | N > 1)$ can be defined as

$$\mathbb{P}(\textit{SINR} > \tau | N > 1) = \sum_{i=2}^{\infty} \mathbb{P}(N = i | N > 1) \mathbb{P}(\textit{SINR} > \tau | N = i) \tag{5.45a}$$

$$= \sum_{i=2}^{\infty} \left(\mathbb{E}[s](1 - \mathbb{E}[s])^{i-2} \int_{x=0}^{\infty} \exp\left(\frac{-\pi x^2 \mathbb{E}[s] \lambda_1}{\rho(\tau, \alpha)^{-1}} + \frac{-x^{\alpha} \sigma_N^2 \tau}{P} \right) g_i dx \right) + \int_{x=0}^{\infty} \exp\left(\frac{-\pi x^2 \mathbb{E}[s] \lambda_1}{\rho(\tau, \alpha)^{-1}} + \frac{-x^{\alpha} \sigma_N^2 \tau}{P} \right) \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2} \right) dx. \tag{5.45b}$$

Proof: See Appendix C.

Figure 5.14 clearly shows that the analytical curves match the numerical results obtained through simulation. For simplicity, a interference-limited scenario, i.e., $\sigma_N^2 = 0$, is considered.

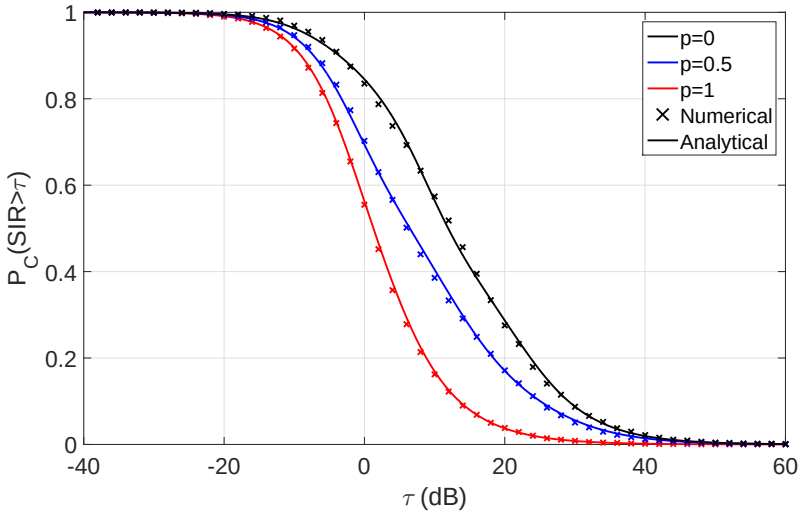


Figure 5.14: Probability of coverage when a load-based algorithm is employed for $\sigma_N^2 = 0$ and $p = 0, 0.5$ and 1.

5.4.4.3 Strategic Sleeping 2: Balancing the frequency resource allocation - A Greedy Algorithm based Strategy

Compared with the load-aware algorithm presented above where only the load at each micro BS was considered, in this section, an heuristic greedy algorithm to provide a fair distribution of the resources among the users and improve the throughput of the network is proposed. Since increasing the throughput of a network subject to a power consumption constraint is of interest, an optimization problem is formulated as follows

$$\begin{aligned} &\min_{||\chi||,P} \sum_{b\in\mathcal{B}} \eta_{ECR}(b) \\ &\text{subject to } \; P_{se} \leq P_{cons} \end{aligned} \tag{5.46}$$

where $\eta_{ECR}(b)$ is the sum ECR value based on equation (5.20) that can be achieved by the set of BSs, $\mathcal{B}$, and P_{se} is the power consumption of the network once some BSs have been switched-off. The optimal solution for energy saving that considers self-backhauling and relaying can be found by exhaustive search. Moreover, as pointed out by [179] and the references therein, convex optimization-based algorithms and their relaxations are computationally expensive for large-scale networks. Thus, the proposed heuristic greedy algorithm is employed to activate in each step the micro BS that minimizes the energy consumption ratio of the system as shown in Algorithm 4. The greedy switching-off algorithm starts with the empty set and iteratively

adds one micro BS location at each step which has the highest energy saving among the set of candidate locations until the saved energy reaches a target value [19].

Algorithm 4 Strategic sleeping 2: Greedy algorithm

1: Initialize $\mathcal{B} = \emptyset, \mathcal{B}_c = BS \in \Phi_k^{(BS)}$, $P_{se} = 0$
2: Set $\eta_{ECR}(\mathcal{B}) = 0$
3: **while** $P_{se} < P_{cons}$ **do**
4: $\quad b = \arg\min_{b \in \mathcal{B}_c} \sum_{b \in \mathcal{B}} (\eta_{ECR}(\mathcal{B} \cup b) - \eta_{ECR}(\mathcal{B}))$
5: $\quad \mathcal{B} \leftarrow \mathcal{B} \cup b$
6: $\quad \mathcal{B}_c \leftarrow \mathcal{B}_c \backslash b$
7: $\quad P_{se} = P_{se} + P_{cons}(b)$
8: **end while**

A greedy algorithm guarantees an optimal solution if i) $\eta_{ECR}(\emptyset) = 0$, ii) $\eta_{ECR}(\mathcal{B})$ is a non-decreasing function, and iii) $\eta_{ECR}(\mathcal{B})$ is submodular [153,154]. In [154], it is shown that under a constant area and bandwidth the throughput satisfies these three properties. In order to show that the greedy algorithm is still a viable solution for an inhomogeneous traffic distribution, the probability that the submodularity condition holds for

$$\eta_{ECR}(\mathcal{B} \cup \{b\}) - \eta_{ECR}(\mathcal{B}) \geq \eta_{ECR}(\mathcal{B}' \cup \{b\}) - \eta_{ECR}(\mathcal{B}'), \tag{5.47}$$

is investigated. The scatter plot in Figure 5.15, shows that about 5% of the cases do not satisfy the inequality. However, the throughput reductions in those cases are relatively small and more than 95% of cases still meet the condition in equation (5.47). Thus, it can be considered that the greedy algorithm should improve the performance of the other two sleeping algorithms, and later in Section 5.5.2, simulation results that support this result are provided.

From the discussion above, it can be pointed out that the greedy switching-off algorithm is the only one among the three algorithms that can take into account the restriction in backhaul links for balancing the network resource allocation and thus achieving an improvement in the system's throughput.

5.5 Numerical Results

In this section, the analytical and numerical results are compared in terms of energy efficiency for the base station deployment and sleeping strategies described in Sections 5.3 and 5.4, respectively.

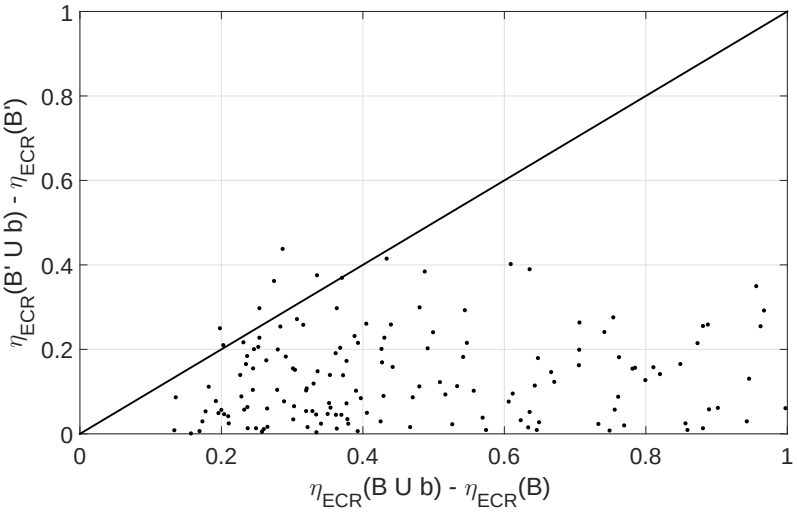


Figure 5.15: Submodularity test under an inhomogeneous traffic distribution.

5.5.1 Base Station Deployment

This subsection presents the energy efficiency analysis for the mm-wave Massive MIMO based wireless backhaul considering the deployment and dynamic operation algorithms presented in Section 5.3. Unless stated otherwise, the results shown in this section pertain to a MIMO system operating at 28 GHz. A narrow band block-fading propagation channel described in equation (2.28) is considered. Moreover, a DF protocol is implemented as the cooperative technique between BSs. For simplicity, the precoder and combiner are assumed to be optimal and $N_s = 5$ data streams are used for transmission. Other important parameters for the simulation can be found in Table 5.1.

Description	Parameter	Value
Transmit power of the micro BS	P	30 dBm
Path Loss exponent	α	Access: LoS = 2.0, NLoS = 3.3
		Backhaul: LoS = 2.0, NLoS = 3.5
Carrier frequency	v	28 GHz
Noise power	σ_N^2	-174 dBm/Hz + $10\log_{10}(W)$ +10
Number of data streams	N_s	5
Number of transmit antennas	N_t	20

Table 5.1: Simulation Parameters of Section 5.5.1.

Firstly, the benefit of considering additional micro BSs within the area of interest is presented. From Figure 5.16, it can be seen that as the number of micro BSs added to the network increases so does the SE. Meanwhile, the EE performance is limited by the additional power consumption incurred by the deployment of micro BSs, so for the sake of comparison, the results obtained

with the optimal solution are also presented. Throughout this section, the optimal solution for the optimization problems described in Algorithms 2 and 3 is computed through an exhaustive search.

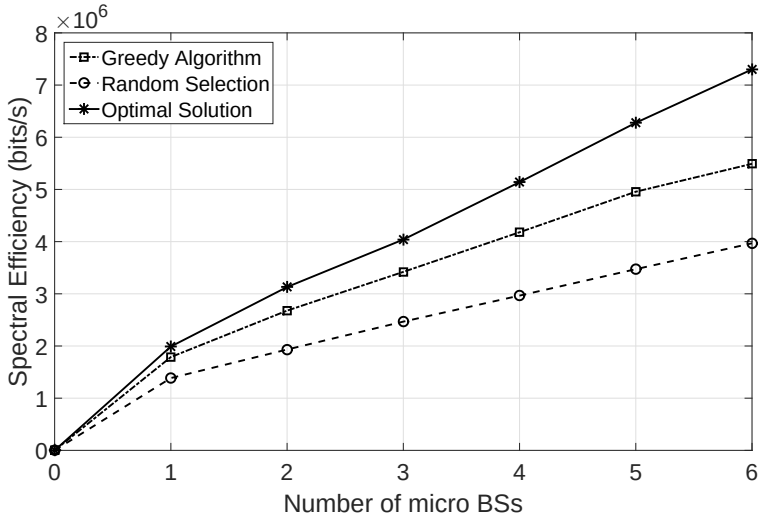
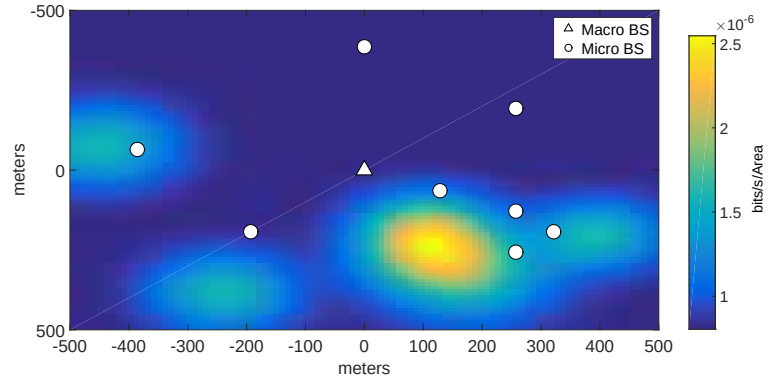


Figure 5.16: Achievable spectral efficiency as a function the number of additional micro BS deployed.

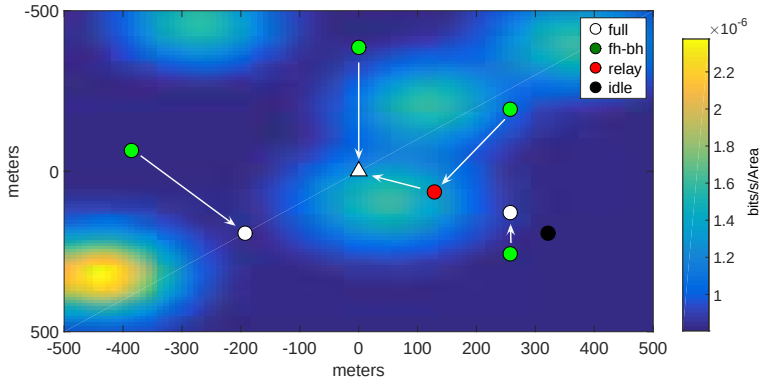
In order to obtain practical results, the traffic profile shown in Figure 5.6 is further considered. As can be seen, low traffic load conditions (lower than 20% of the peak-time) occur for about 67% of the day for a micro BS. Thus, during this period a dynamic operation can be implemented to reduce the energy consumption of the network.

Figure 5.17(a) shows the deployment of the additional micro BSs to support the traffic load at peak-time (12 - 14hrs and 15 - 20 hrs) obtained through Algorithm 3. Meanwhile, Figure 5.17(b) presents the reconfiguration and wireless cooperation carried out by the BSs for a low load period. As it is noted, as the traffic load and distribution of the users vary, the tasks performed by each of the BS may change as well. The only exception is the macro BS, which is considered to be active the whole time to carry out scheduling, resource allocation and support for high-mobility users. In particular, it can be noticed that in areas where the traffic load is low the BSs adopt relaying tasks or are switched-off. In both cases, the EE of the network is improved by reducing the total power consumption. Specially, the EE performance is improved when a cooperative system is implemented as the distances between BSs are scaled down and therefore, the high path loss that mm-wave frequencies experience is reduced.

Figure 5.18 shows the average number of BSs that are active regarding the time of the day.



(a)



(b)

Figure 5.17: (a) Full operation mode of the additional micro BSs during peak-time traffic load (2pm) and (b) dynamic operation mode of each micro BSs when the traffic load reduces (4pm). The arrows indicate the association between BSs when Algorithm 3 is used to dynamically change the operating mode of the BSs.

During peak-time traffic most of the BSs remain active to meet the traffic demand. When the load of the network reduces, the dynamic algorithm switches-off some of the BSs to save energy. The results obtained through Algorithms 2 and 3 are compared with a random selection, where the BSs are deployed without taking into account traffic profiles and deployment costs. From Figure 5.18, it can be seen that, in general, the number of active antennas are lower for the greedy algorithm. Furthermore, on average a reduction of 17.6% in terms of the number of BS required to satisfy the demand at peak-time is achieved, compared to random allocation.

Considering that it is of particular interest to analyse the energy efficiency of the system for dynamic operation, Figures 5.19(a) and 5.19(b) show the operation modes adopted by the micro BSs regarding the traffic profile. As can be seen, when the demand in data rate reduces, some of the micro BSs operating in fronthaul-backhaul mode are instead used as a relay or switched-

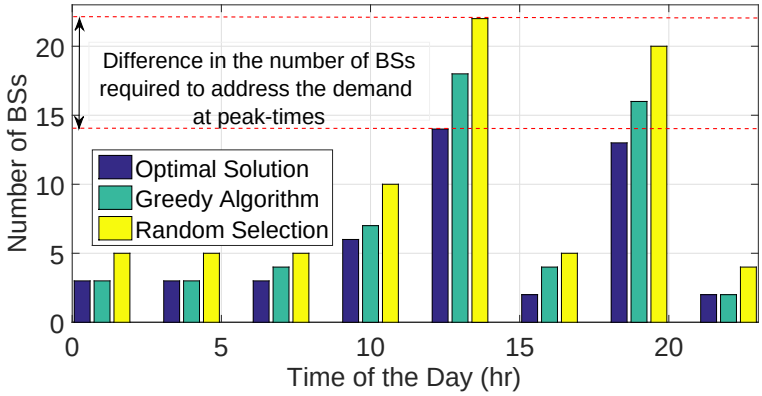
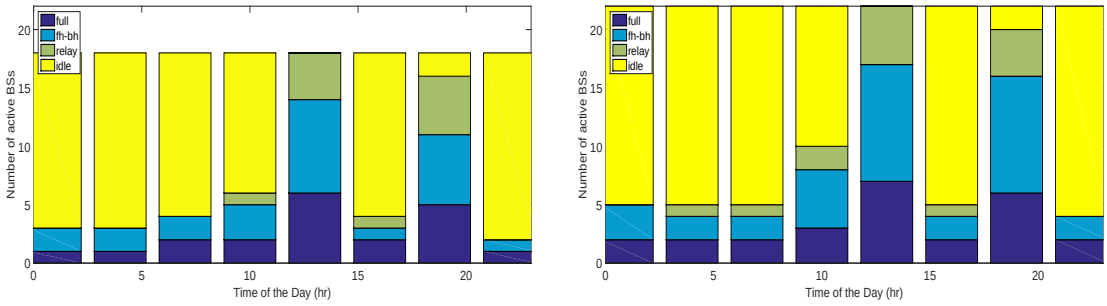


Figure 5.18: Number of active BSs regarding the time of the day.

off. Furthermore, it can also be noticed that the number of BSs required to support the traffic demand during different times of the days follows the results shown in Figure 5.18, as expected.



(a) (b)

Figure 5.19: Operation mode when a (a) greedy algorithm and (b) random selection are used for deploying the additional BSs.

As a result, in Figure 5.20, it can be seen that the EE of the system considerably improves when the dynamic operation algorithm is considered, achieving a maximum value of 0.34 compared with the 0.22 bits/s/Hz for the random selection. This is can be explained considering the fact that the traffic aware algorithm can better determine the number of active BSs required to satisfy the current demand for data rate. Compared with a system without dynamic operation, a significant reduction in the power consumption is achieved when the system is under-utilized during the low traffic load periods.

5.5.2 Sleeping Strategies

In this section, the numerical and analytical results for the random, load-based and greedy sleeping policies described in Section 5.4 are compared. Unless stated otherwise, the results

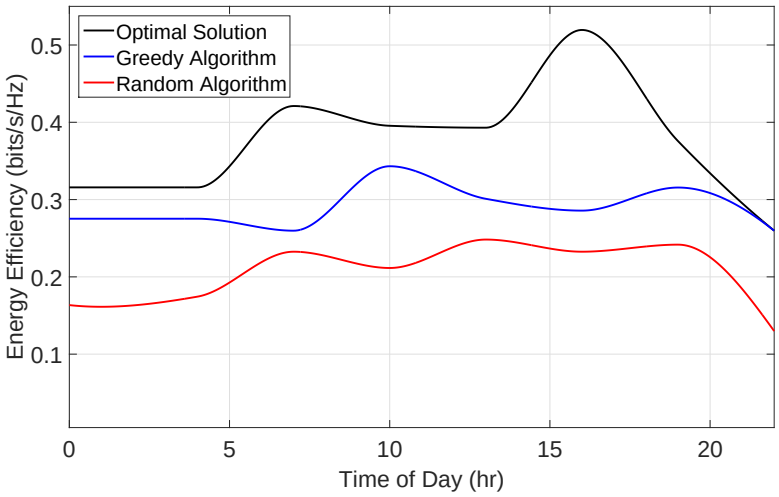


Figure 5.20: Energy consumption considering the active/inactive BSs for a specific traffic distribution.

shown in this subsection pertain to a mm-wave system operating at 73 GHz. A narrow band block-fading propagation channel is considered. Moreover, a DF protocol is implemented as the cooperative technique between BS and its tagged A-BS. Other important simulation parameters can be found in Table 5.2.

Description	Parameter	Value
Fraction of anchored (wired) BSs	w	0.3
BS PPP density	λ_1	50 per km^2
UE PPP density	$\lambda^{(UE)}$	1000 per km^2
Variance of the Thomas cluster process	σ^2	100 m
mm-wave bandwidth	W	2 GHz
Transmit power of micro BS	P	30 dBm
Path Loss exponent	α	Access: LoS = 2.0, NLoS = 3.3 Backhaul: LoS = 2.0, NLoS = 3.5
Carrier frequency	v	73 GHz
Antenna gain at the BS	G_B	18 dB
Noise power	σ_N^2	-174 dBm/Hz + $10\log_{10}(W)$ +10

Table 5.2: Simulation Parameters of Section 5.5.2.

Figure 5.21 shows the probability of coverage for different values of the cluster variance σ^2 and a fixed value $p = 0.5$, which implies that half of the traffic of the network is not correlated with any BS. From Figure 5.21 and following the discussion in Section 5.4.3 is evident that reducing the cluster variance reduces the distance between the typical user and its tagged BS, thus the probability of coverage is improved. Moreover, the plot clearly indicates that the analytical curves match the numerical results obtained through simulation.

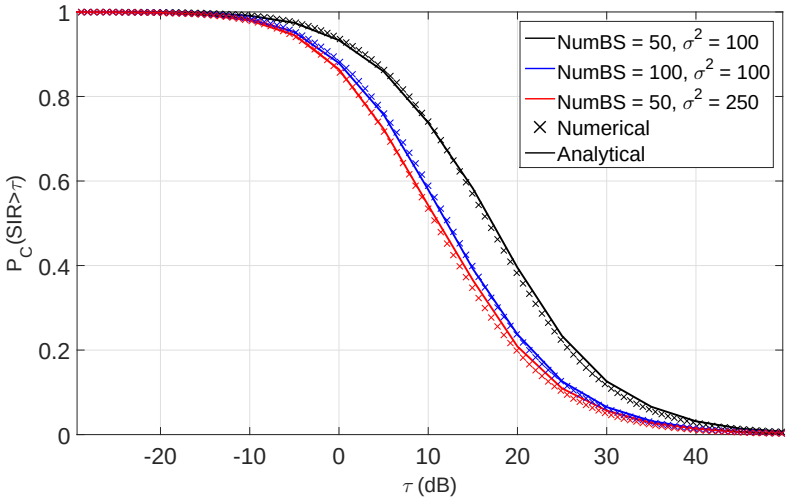
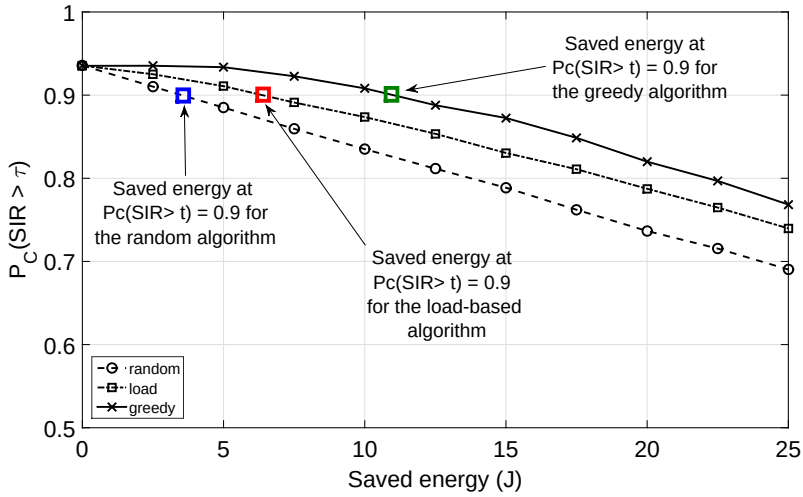


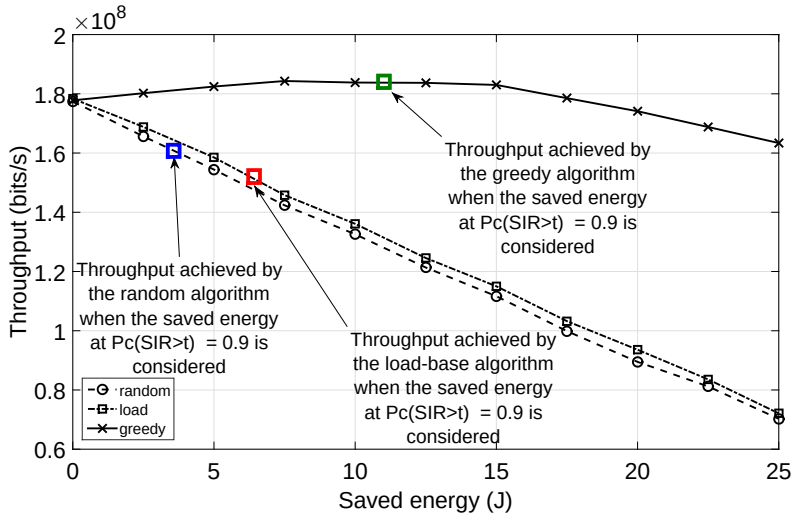
Figure 5.21: Probability of coverage for different density of BSs and cluster variance, σ^2.

Continuing with the analysis of the system's performance, in the following, the trade-off between the probability of coverage and throughput when sleeping strategies are employed to reduce the power consumption of HetHetNets is analysed. Figure 5.22(a) compares the probability of coverage for the sleeping policies discussed in Section 5.4 when the inhomogeneity of the network is set to $p = 0$, i.e., all the users are correlated with the BS. It is shown that the random algorithm performs the worst when switching-off BSs, showing a gap of about 9J in saved energy with respect to the greedy algorithm (at $\mathbb{P}_C = 0.9$). It is important to note that in a highly clustered network, the greedy algorithm performs 5J better than the load-based algorithm since switching-off BSs based on the load inevitably increases the average distance between the typical user and its tagged BS. Comparing these results with those in Figure 5.23(a), it can be noticed that when the superposition of the user distributions, i.e., $p = 0.5$, is considered the load-based algorithm outperforms the greedy algorithm by about 13J. The reason for this behaviour is two-fold, first, the cost function of the greedy algorithm prioritizes the throughput and not the probability of coverage and second, in highly clustered networks, the serving distance is reduced.

Figure 5.22(b) shows the total throughput of the HetHetNet when $p = 0$. In order to fairly compare the performance of the sleeping polices, the probability of coverage is set at 90% for each algorithm. Blue, red and green squares are used to indicate the throughput of the system when the random, load-based and greedy algorithms are employed, respectively. Based on the results in Figure 5.22(b), it can be noticed that the greedy algorithm considerably outperforms



(a)



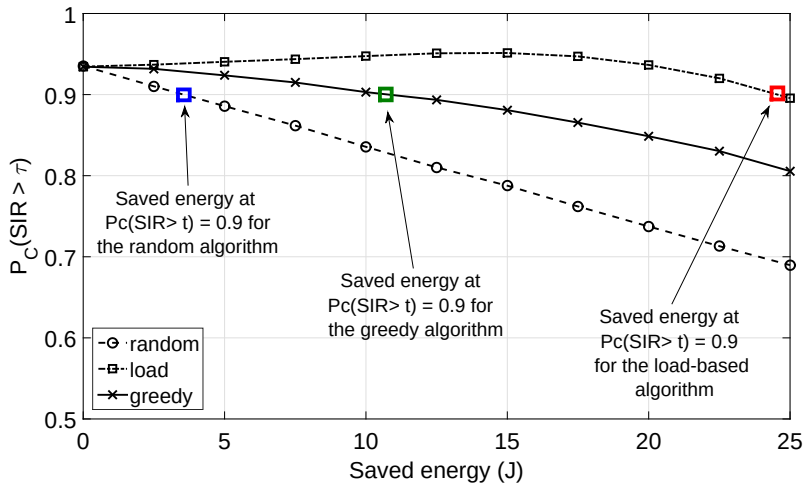
(b)

Figure 5.22: (a) Probability of coverage and (b) throughput vs. saved energy for a HetHetNet with NumBS = 50, NumUEs = 1000, $p = 0$ and τ = 0 dB. The squares indicate the saved energy when the probability of coverage is at 90%.

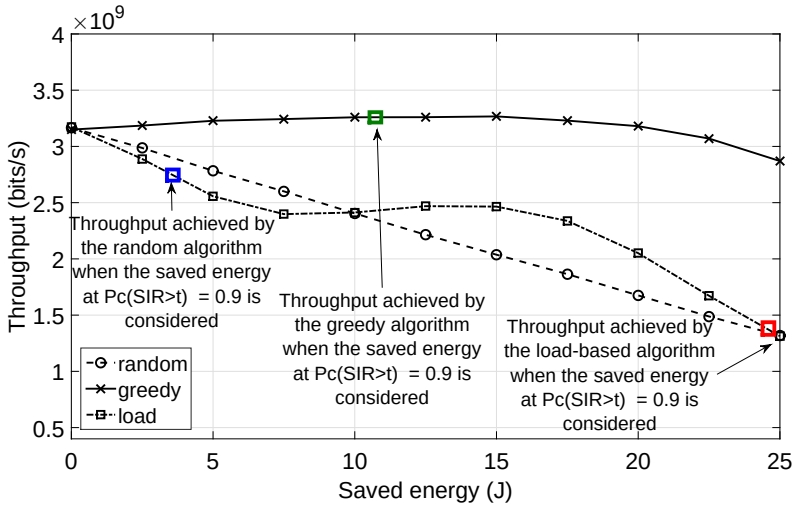
the other sleeping polices for a data rate of 150 MB/s. From Figures 5.22(a) and 5.22(b), it can be pointed out that even though the load-based algorithm shows improvement in terms of saved energy with respect to the random algorithm, the trade-off in this case is seen in the overall throughput, since the random policy provides a higher throughput when a fixed probability of coverage is considered.

In order to understand how the throughput of the system is affected when independent users

are considered in the system, p is set to 0.5 ($p = 0.5$). Figure 5.23(b) shows that in terms of throughput the greedy algorithm achieves the best trade-off of energy saving and data throughput for a wide range of scenarios, including highly homogeneous PPP and highly heterogeneous clustered wireless networks. From this plot, it can also be noticed that the throughput obtained by the load-based algorithm is very affected by the presence of A-BS base stations that limit the available bandwidth per user and by the reduction of the interference when the BSs are switched-off.



(a)



(b)

Figure 5.23: (a) Probability of coverage and (b) throughput vs. saved energy for a HetHetNet with NumBS = 50, NumUEs = 1000, $p = 0.5$ and τ = 0 dB. The squares indicate the saved energy when the probability of coverage is at 90%.

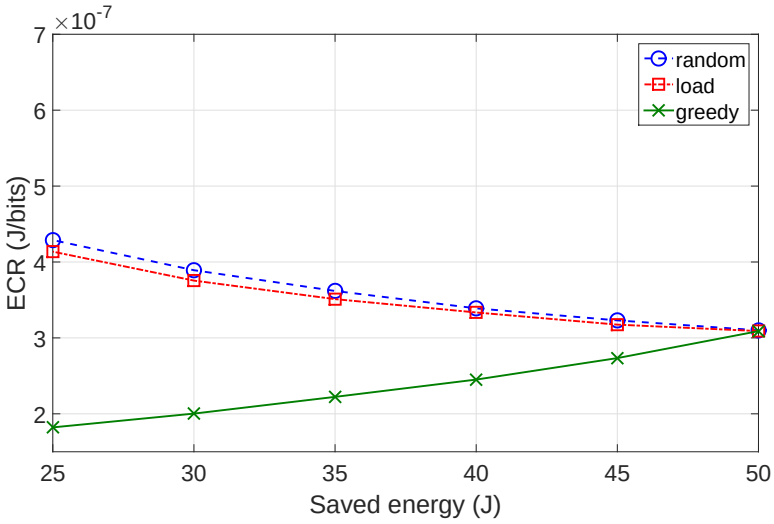
NumBS = 50, p = 0, UEs = 1000				
$\mathbb{P}_C$	Marker colour	Algorithm	Energy (J)	Throughput (bits/s)
0.90	blue	R	3.5	0.16e9
	red	L	6.3	0.15e9
	green	G	10.8	0.18e9
NumBS = 50, p = 0.5, UEs = 1000				
$\mathbb{P}_C$	Marker	Algorithm	Energy (J)	Throughput (bits/s)
0.90	blue	R	3.6	2.92e9
	red	L	24	1.39e9
	green	G	10.5	3.24e9
R = random, L = load-based, G = greedy algorithm				

Table 5.3: Comparison of the saved energy and throughput when the system is highly clustered and the superposition of cluster and homogeneous processes is considered.

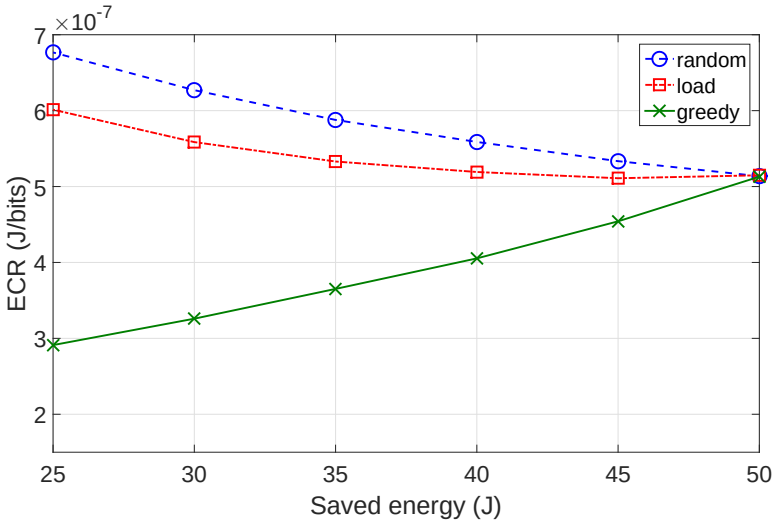
Finally, the ECR performance of the system is analysed. Figures 5.24(a) and 5.24(b) show the ECR achieved by the sleeping polices when $p = 0$ and $p = 0.5$, respectively. From these plots, it can be seen that the best performance is achieved by the greedy algorithm having the lowest ECR value for the whole range of saved energy intervals. Moreover, a better performance from all the sleeping policies can be achieved in a highly clustered network due to the throughput depending on the serving distance. On the other hand, it can be noticed the ECR value of the load-algorithm shows a heavy dependence on the interference and the fraction of A-BSs available according to the throughput results in Figures 5.22 and 5.23.

From Figure 5.24(a), it can also be noticed that the ECR performance of all sleeping policies tends to the same value as the number of BSs that are switched-off increases, i.e., there is a saturation in throughput that can be achieved by the greedy algorithm and thus, continuing to switch BSs off increases the power consumed to carry out the cooperation between the BSs and the A-BSs. Hence, after this crossing point, the advantage provided by the greedy algorithm in terms of ECR is negligible compared with the other two algorithms.

In general, the more clustered the network, the smaller the serving distance between the typical user and its tagged BS and thus, the better the performance in terms of probability of coverage and throughput. For the sake of clarity, a summary of the saved energy using the three sleeping policies is presented in Table 5.3, where the target probability of coverage is set to 0.9 and the blue, red and green squares in Figures 5.22 and 5.23 are the corresponding markers for the random, load-based and greedy algorithm, respectively.



(a)



(b)

Figure 5.24: Energy consumption ratio (ECR) with NumBS = 50, NumUEs = 1000, τ = 0 dB and (a) p = 0 and (b) p = 0.5, respectively.

5.6 Summary

In this chapter, a millimetre-wave system to enable self-backhauling in future wireless networks was considered. Aiming to handle the inhomogeneous distribution of increasing traffic demand over time and space in 5G wireless communications systems, a tunable model that can cover a wide range of scenarios from a homogeneous PPP to highly heterogeneous clustered wireless networks was presented, where a Thomas cluster process is employed to generate the clustered

users. Moreover, energy-efficient base station deployment and traffic aware dynamic operation algorithms for time-varying scenarios were presented. Furthermore, the distribution of the serving distance was analysed and the probability of coverage and throughput when the wireless self-backhaul system with constrained frequency resources is employed were provided.

Numerical results showed that the average number of additional micro BS required to support the traffic load at peak-time was 17.6% lower compared with a random selection. Additionally, it was shown that the EE performance of the network can be improved from 0.22 to 0.34 bits/s/Hz when the dynamic operation algorithm is considered.

Finally, aiming to optimize the energy efficiency of the system, the use of energy-efficient sleeping algorithms was investigated. Three different policies for switching-off base stations were considered: random, load-based and greedy algorithms. The probability of coverage for the random and load-based sleeping policies was derived and the performance of the heuristic greedy algorithm studied.

Numerical results showed that highly clustered networks exhibit a smaller average serving distance and thus a better probability of coverage. Moreover, the greedy algorithm provided a higher throughput since the load-based and random algorithms do not consider limitations in frequency resources for backhauling while switching-off base stations. Nevertheless, the load-based algorithm outperformed the greedy algorithm in terms of probability of coverage in moderately clustered scenarios since the overall throughput of the network is prioritized by the cost function of the greedy algorithm. The load-based algorithm attained energy savings of 13.5 J higher than the greedy algorithm, but the greedy algorithm achieved 1.85 Gb/s higher overall throughput compared with the load-base algorithm.

Chapter 6
Conclusions and Future Work

This thesis has contributed to the design of energy-efficient cooperative wireless systems when a large number of antennas at the transmitter is considered. In this chapter, Section 6.1 will emphasize the key contributions of the thesis. Meanwhile, some limitations of the work and suggestions for future work will be discussed in Section 6.2.

6.1 Conclusions

This thesis extended the application of cooperative communications to MIMO wireless systems equipped with large transmit antenna arrays. Focused on providing insights for energy-efficient systems, different approaches were considered such as: designing low complexity quantisation techniques, deploying and operating a network for energy-efficient performance and studying self-backhauling mm-wave systems. The key contributions of this thesis are summarised as follows.

6.1.1 Low Complexity Design

When the number of antennas employed for transmission increases from tens to hundreds different challenges are imposed for the system's implementation. One of them related with the computational complexity required to carried out the cooperation between terminals. Considering that the codebook design plays a key role in the system's complexity and its implementation, when a CF protocol is used for wireless cooperation, then the design of low complexity and reliable codebooks is required to enable future 5G systems. In this work, low complexity quantisation techniques were proposed. Numerical simulations suggest that a comparable performance with the optimum vector quantiser can be achieved when the probability distribution function of the signal is considered during the codebook design. In particular, if the signals follows a Gaussian distribution the granular and overload errors should be taken into account before quantisation. The reduced complexity of the scalar techniques presented is particularly

valuable and attractive for practical wireless communication systems, and enables scaling up the benefits of cooperative MIMO systems.

6.1.2 Massive Cooperative MIMO systems

Despite the benefits provided from multi-antenna cooperative systems, this comes at the expense of linearly increasing the power consumption at the transmitter due to the use of multiple radio frequency chains to drive the antennas. Thus, in this work, the trade-off between capacity and total power consumption was evaluated. Numerical results suggest that not only a practical trade-off between capacity and power consumption can be obtained but also that Massive cooperative MIMO systems are suitable for practical implementation when limited information of the channel is available at the transmitter, making these systems robust not only to channel estimation errors but also to imperfections of the circuit components. Additionally, it was shown that even though the design of power control schemes is in general not a trivial task, the degrees of freedom in Massive MIMO systems allows using simple linear precoding schemes, greatly simplifying its design. Numerical results suggest that the implementation of power allocation algorithms are more significant when the ratio between the number of transmit and receive antennas approaches to the unit.

6.1.3 Energy Efficient Backhaul systems

Compared with 4G systems where the locations of both users and base stations were modelled as homogeneous Poisson point processes, due to the nature of 5G deployments, the users are more likely to appear clustered towards the base stations, rather than being uniformly distributed around them. In this work, a millimetre-wave system to enable self-backhauling for inhomogeneous wireless networks was considered. Numerical results suggest that mm-wave can be considered as an enabler for 5G systems, since highly clustered networks exhibit a smaller average serving distance and thus mm-wave frequencies can overcome attenuations due the propagation. Moreover, the use of large number of antennas at the transmitter supported and enhanced by the mm-wave spectrum provides a solution to the backhaul bottleneck, enabling the adoption of self-backhauling systems. Finally, numerical results show that policies for switching-off base stations can be considered as an alternative to improve the energy efficiency of the system when a backhaul network with limited frequency resources is considered.

6.2 Future Work

This work has identified promising technologies for energy-efficient 5G systems using cooperative wireless communications. Nevertheless, the proposed schemes have some limitations due to the initial modelling assumptions. Thus, there are several research directions that can be extended. Some suggestions are listed below:

- In Chapter 3, low complexity quantisation algorithms were presented. Simulations results showed that the distribution of the signals plays a key role in the performance of the quantiser. Even though the Lloyd-Max algorithm takes into account the distribution of the signals during the codebook design, a further improvement in the trade-off between complexity and performance can be achieved if a two-stage quantiser is considered [180]. Additionally, new approaches to improve the performance of the compress-and-forward protocol such as: bottleneck graphs, bi-section method and lattices can be considered in cooperative wireless systems with large antenna arrays [181].

- In Chapter 4, multiple relays assist the destination node to increase its throughput via wireless cooperation. Nevertheless, it was assumed that the destination and relays nodes share their data through error free links. Since in large scale wireless networks, cooperation can lead to severe interference and thus higher packet loss and throughput reduction, the results of this chapter can be improved by incorporating interference-aware wireless cooperation algorithms. System level simulations are required to evaluate the trade-off between alleviating co-channel interference and exploiting cooperative diversity in large antenna arrays setups.

- Even though the gap between 5G backhaul requirements and backhaul capabilities has been significantly reduced with the use of mm-waves systems. In Chapter 5, it was assumed that the backhaul network was only restricted in terms of total power consumption and capacity. Thus, the results presented can be further improved if the modelling of inhomogeneous HetNets includes disparities in terms of synchronisation and latency. Enabling the 5G backhaul solutions to exploit the heterogeneity of the network while offering the quality of service required by the users.

- In Chapter 5, the design of an energy-efficient self-backhauling mm-wave cooperative system was proposed. Nevertheless, it was assumed that perfect channel state information was available at the transmitter and that digital baseband beamforming was carried

out at the base station before transmission. Unfortunately, perfect channel state information is difficult to achieve due to the large number of antennas and the small signal-to-noise ratio before beamforming. Moreover, the high cost of mixed signal components makes digital baseband beamforming an impractical solution. Thus, an extended analysis of the system's performance when channel estimation algorithms and hybrid precoding solutions tailored to mm-wave frequencies are considered could enriched the modelling and provide more insights into the application of mm-wave systems for 5G deployments.

Appendix A
Ratio of the distances to the tagged BS and strongest interferer

With a interference-limited network, the noise power can be considered as negligible. The SIR, γ, in dB can be calculated as $\gamma = P_S - I_{agg}$, where P_S is the transmit power associated with the tagged A-BS or BS and I_{agg} the aggregate interference. Thus, the SIR considering only the strongest interferer can be upper-bounded as follows

$$\gamma = P_S - I_{agg} \leq \hat{\gamma} = P_S - P_I \tag{A.1a}$$
$$= P + G - PL_S + h_S - (P + G - PL_I + h_I) \tag{A.1b}$$
$$= PL_S - PL_I + (h_S - h_I) \tag{A.1c}$$

where $PL_i[dB] = 10\alpha \log_{10}(R_i)$ with $i \in \{S, I\}$ is the path loss , $\alpha > 2$, $h_S \sim \mathcal{N}(\mu, \sigma_N^2)$ and $h_I \sim \mathcal{N}(\mu, \sigma_N^2)$. Thus, the upper-bound $\hat{\gamma}$ can be rewritten as

$$\hat{\gamma} = 10\log_{10}(R_S^\alpha) - 10\log_{10}(R_I^\alpha) + h \tag{A.2a}$$
$$\hat{\gamma} = -\frac{10}{\log(10)} \cdot \alpha \cdot \log\left(\frac{R_S}{R_I}\right) + h \tag{A.2b}$$
$$\hat{\gamma} = -\frac{10}{\log(10)} \cdot \alpha \cdot \log(R_{S,I}) + h \tag{A.2c}$$

where $0 \leq R_S \leq R_I$ and $R_{S,I} \in [0, 1]$ and $h = h_S + h_I \sim \mathcal{N}(2\mu, 2\sigma_N^2)$. Considering the PDF distribution of the serving distance and strongest interferer, the CDF of the ratio of the distances to the tagged BS and strongest interferer, $F_r(R_{S,I})$, is given by

$$F_r(R_{S,I}) = P(r \leq R_{S,I}) = (1-p)\int_0^{\infty}\int_0^{r_2 R_{S,I}} f_{R_0}(r_1, r_2)dr_1 dr_2 + p\int_0^{\infty}\int_0^{r_2 R} f_{R_1}(r_1, r_2)dr_1 dr_2 \tag{A.3a}$$

$$= (1-p)\int_0^{\infty}\int_0^{r_2 R_{S,I}} \left(2\pi\lambda_1 r_2 e^{-\pi(r_2^2 - r_1^2)}\right)\left(\frac{r_1}{\sigma^2} e^{\frac{-r_1^2}{2\sigma^2}}\right) dr_1 dr_2 + pR_{S,I}^2 \tag{A.3b}$$

$$= (1-p)\frac{R_{S,I}^2}{R_{S,I}^2 - 2\pi\lambda_1(R_{S,I}^2 - 1)\sigma^2} + pR_{S,I}^2. \tag{A.3c}$$

where equation (A.3a) accounts for the closest BS and strongest interferer in each of the tiers. Taking the derivative of $F_r(R_{S,I})$, the PDF of the ratio of the distances, $f_r(R_{S,I})$, can be obtained as follows

$$f_r(R_{S,I}) = (1-p)\left(-\frac{R_{S,I}^2(2R_{S,I} - 4\pi\lambda_1 R_{S,I}\sigma^2)}{(R_{S,I}^2 - 2\pi\lambda_1(R_{S,I}^2 - 1)\sigma^2)^2} + \frac{2R_{S,I}}{R_{S,I}^2 - 2\pi\lambda_1(R_{S,I}^2 - 1)\sigma^2}\right) + 2pR_{S,I}, \qquad 0 \leq R_{S,I} \leq 1. \tag{A.4a}$$

Finally, using the inverse transform theorem, the PDF of the logarithm of $R_{S,I}$, i.e., $Y = \log(R_{S,I})$, is obtained as

$$f_Y(y) = e^y(1-p)\left(-\frac{e^{2y}(2e^y - 4\pi\lambda_1 e^y\sigma^2)}{(e^{2y} - 2\pi\lambda_1(e^{2y} - 1)\sigma^2)^2} + \frac{2e^y}{e^{2y} - 2\pi\lambda_1(e^{2y} - 1)\sigma^2}\right) + 2pe^{2y}, \qquad y \leq 0. \tag{A.5a}$$

Appendix B
Probability of Coverage with Random Sleeping

In order to account for the background noise, σ_N^2, its contribution is incorporated to the probability of coverage (equation (5.32)) by multiplying its Laplace transform as follows

$$\mathbb{P}_{C_H}^{\sigma_N} = \int_{r>0} 2\pi q\lambda_1 r \exp(-\pi q\lambda_1(\rho(\tau,\alpha)+1)r^2)\exp(\tau r^\alpha \sigma_N^2 P^{-1})dr, \tag{B.1}$$

$$\mathbb{P}_{C0}^{\sigma_N} = \int_{r>0} 2\pi q\lambda_0 r \exp(-\pi q\lambda_0(\rho(\tau,\alpha)+1)r^2)\exp(\tau r^\alpha \sigma_N^2 P^{-1})dr, \tag{B.2}$$

and

$$\begin{aligned}\mathbb{P}_{C1}^{\sigma_N} = 2\pi q\lambda_1 \int_0^\infty \int_{x_i}^\infty \left[\frac{1}{\sigma^2}\frac{r}{1+\tau(r/x_i)^{-\alpha}}\exp\left(\frac{-r^2+x_i^2}{2\sigma^2}\right)dr\right] \\ \times\, x\cdot\exp\left(-\pi x^2(1+\rho(\tau,\alpha))q\lambda_1+\lambda_0)\right)dx.\end{aligned} \tag{B.3a}$$

Substituting equations (B.1), (B.2) and (B.3a) into (5.32), the probability of coverage of the random sleeping algorithm is given by

$$\mathbb{P}_{C_{RS}} = (1-p)(\mathbb{P}_{C0}^{\sigma_N}+\mathbb{P}_{C1}^{\sigma_N})+p\mathbb{P}_{C_H}^{\sigma_N}. \tag{B.4}$$

If the background noise, $\sigma_N^2 = 0$, is not considered, then equation (B.4) reduces to

$$\mathbb{P}_{C_{RS}} = (1-p)\left(q\left(\mathbb{P}_{C0}+\mathbb{P}_{C1}\right)+\frac{1-q}{1+G}\right)+p\frac{1}{1+G}. \tag{B.5}$$

Appendix C
Probability of Coverage with the Load-based Algorithm

As suggested in [155], the first step to obtain the probability of coverage is to condition the sleeping algorithm on the activity of a Poisson-Voronoi region $a(x)$. Thus, all the micro BSs are enumerated in increasing order of distance from the typical user, starting from $N = 1$, i.e., the serving BS. Moreover, i.i.d. random variables, $A_i \sim A$, are assigned to each micro BS such that $A \in [0, 1]$, where A_i represents the user activity within the Poisson-Voronoi cell and $f_A(x)$ describes its PDF. From [155], it is known that the probability of coverage of the load-based algorithm can be described by

$$\mathbb{P}_{C_{LB}} = \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x\mathbb{P}(\textit{SINR} > \tau|x) f_A(x) dx \tag{C.1a}$$

$$= \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x\mathbb{P}(N = 1)\mathbb{P}(\textit{SINR} > \tau|N = 1) f_A(x) dx + \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x\mathbb{P}(N > 1)\mathbb{P}(\textit{SINR} > \tau|N > 1) f_A(x) dx \tag{C.1b}$$

$$= \frac{1}{\mathbb{E}[a(x)]} \int_0^1 xs(x)\mathbb{P}(\textit{SINR} > \tau|N = 1) f_A(x) dx + \frac{1}{\mathbb{E}[a(x)]} \int_0^1 x(1 - s(x))\mathbb{P}(\textit{SINR} > \tau|N > 1) f_A(x) dx \tag{C.1c}$$

where (C.1a) is, by definition, the coverage of probability weighted over the active user links, (C.1b) partitions the probability of coverage into the event when the nearest micro BS is awake and the event when it is asleep, (C.1c) uses the Laplace transform of the remaining interferers to represent the probability of coverage. The term $\mathbb{P}(\textit{SINR} > \tau|N = 1)$ corresponds to the PDF presented by equations (5.23a) or (5.22a) if the nearest BS is correlated or not with the user location, respectively. In order to calculate, $\mathbb{P}(\textit{SINR} > \tau|N > 1)$, two cases are considered: if the closest BS to the typical user is (a) its cluster centre or (b) any other micro BS. When case (a) occurs, then the interferers are uniform distributed regarding the typical user location. Nevertheless, when case (b) is considered, the contribution of the cluster centre as an interferer

must be considered as follows,

$$\mathbb{P}(\textit{SINR} > \tau | N > 1) = \sum_{i=2}^{\infty} \mathbb{P}(N = i | N > 1) \mathbb{P}(\textit{SINR} > \tau | N = i) \tag{C.2a}$$

$$= \sum_{i=2}^{\infty} \left[\mathbb{E}[s](1 - \mathbb{E}[s])^{i-2} \int_{x=0}^{\infty} \exp\left(-\pi x^2 \mathbb{E}[s] \lambda_1 \rho(\tau, \alpha) + \frac{-x^{\alpha} \sigma_N^2 \tau}{P} \right) \times g_i dx \right] + \int_{x=0}^{\infty} \frac{x}{\sigma^2} \exp\left(-\pi x^2 \mathbb{E}[s] \lambda_1 \rho(\tau, \alpha) + \frac{-x^{\alpha} \sigma_N^2 \tau}{P} - \frac{x^2}{2\sigma^2} \right) dx \tag{C.2b}$$

where $g_i = 2(\pi\lambda_1)^i x^{2i-1}/(i-1)! \exp(-\pi x^2 \lambda_1)$. Notice that the last term in (C.2b) should be discarded if the cluster centre is switched-off as a result of the sleeping strategy adopted and thus, (C.2a) reduces to the results derived in [155].

Appendix D
Publication List

Journal Papers

1. Miryam Gonzalez-Perez and J. Thompson, *"Energy Efficient Inhomogeneous Cellular Networks,"* Submitted to *IEEE Transactions on Communications*.

Conference Papers

1. M. Gonzalez and J. Thompson, *"An energy efficient base station deployment for mm-wave based wireless backhaul*," 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, 2016, pp. 1-6.

2. M. G. Gonzalez-Perez and J. Thompson, *"Analysis of the energy efficiency of a Virtual MIMO system with a large antenna array*," IEEE EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), Salamanca, 2015, pp. 1-6.

3. M. G. Gonzalez-Perez and J. Thompson, *"A Low Complexity Quantization Technique for Virtual MIMO Systems*," 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, 2015, pp. 1-5.

References

[1] C. T. K. Ng, I. Maric, A. J. Goldsmith, S. Shamai, and R. D. Yates, "Iterative and one-shot conferencing in relay channels," in *2006 IEEE Information Theory Workshop - ITW '06 Punta del Este*, pp. 193–197, March 2006.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, pp. 101–107, June 2011.

[3] O. Boswarva, "Sitefinder mobile phone base station database." `http://datashare.is.ed.ac.uk/handle/10283/2626`, 2012. Accessed: 2017-08-17.

[4] A. Abrol and R. K. Jha, "Power optimization in 5G networks: A step towards green communication," *IEEE Access*, vol. 4, pp. 1355–1374, 2016.

[5] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1164–1179, June 2014.

[6] "Ericsson mobility report," tech. rep., Ericsson, 2017. Accessed: 2018-01-16.

[7] E. Bjrnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 3059–3075, June 2015.

[8] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1065–1082, June 2014.

[9] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–7, May 2015.

[10] Alcatel-Lucent, "Alcatel-lucent demonstrates up to 27 percent power consumption reduction on base stations deployed by china mobile: Software upgrades can offer exceptional power and cost savings for mobile operators worldwide," tech. rep., February 2009.

[11] F. Han, Z. Safar, W. S. Lin, Y. Chen, and K. J. R. Liu, "Energy-efficient cellular network operation via base station cooperation," in *2012 IEEE International Conference on Communications (ICC)*, pp. 4374–4378, June 2012.

[12] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, pp. 30–37, June 2011.

[13] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, pp. 3037–3063, Sept 2005.

[14] X. Cai, Y. Yao, and G. B. Giannakis, "Achievable rates in low-power relay links over fading channels," *IEEE Transactions on Communications*, vol. 53, pp. 184–194, Jan 2005.

[15] M. Haenggi, *Stochastic Geometry for Wireless Networks*. New York, NY, USA: Cambridge University Press, 1st ed., 2012.

[16] M. G. Gonzalez-Perez and J. Thompson, "A low complexity quantization technique for virtual MIMO systems," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2015.

[17] M. G. Gonzalez-Perez and J. Thompson, "Analysis of the energy efficiency of a virtual MIMO system with a large antenna array," in *IEEE EUROCON 2015 - International Conference on Computer as a Tool (EUROCON)*, pp. 1–6, Sept 2015.

[18] M. Gonzalez and J. Thompson, "An energy efficient base station deployment for mm-wave based wireless backhaul," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6, Sept 2016.

[19] M. Gonzalez-Perez and J. Thompson, "Energy efficient inhomogeneous cellular networks," *submitted to IEEE Transactions of Green Communications and Networking*, December 2017.

[20] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. New York, NY, USA: Cambridge University Press, 2007.

[21] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, pp. 198–218, Feb 2004.

[22] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[23] E. Telatar, "Capacity of multi-antenna gaussian channels," *Transactions on Emerging Telecommunications Technologies*, vol. 10, no. 6, pp. 585–595, 1999.

[24] J. Jiang, *Virtual-MIMO Systems with Compress-and-Forward Cooperation*. PhD thesis, The University of Edinburgh, 2011.

[25] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.

[26] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[27] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[28] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, Oct 1998.

[29] T. L. Marzetta, "Massive MIMO: an introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[30] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, February 2014.

[31] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive mimo: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 742–758, Oct 2014.

[32] J. Vieira, F. Rusek, and F. Tufvesson, "Reciprocity calibration methods for massive mimo based on antenna coupling," in *2014 IEEE Global Communications Conference*, pp. 3708–3712, Dec 2014.

[33] H. Q. Ngo, *Massive MIMO: Fundamentals and system designs*, vol. 1642. Linköping University Electronic Press, 2015.

[34] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, Mobicom '12, pp. 53–64, ACM, 2012.

[35] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser mimo," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1815–1831, April 2014.

[36] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 172–179, February 2013.

[37] V. Jungnickel, V. Pohl, and C. von Helmolt, "Capacity of MIMO systems with closely spaced antennas," *IEEE Communications Letters*, vol. 7, pp. 361–363, Aug 2003.

[38] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile ad hoc networks," *IEEE Signal Processing Magazine*, vol. 23, pp. 18–29, Sept 2006.

[39] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, Oct 2003.

[40] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1099–1109, Aug 2004.

[41] P. Cao, M. Gonzalez-Perez, and J. S. Thompson, *Relaying in green communication systems*, pp. 207–242. Telecommunications, Institution of Engineering and Technology, 2016.

[42] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, pp. 3062–3080, Dec 2004.

[43] R. U. Nabar and H. Bolcskei, "Space-time signal design for fading relay channels," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 4, pp. 1952–1956, Dec 2003.

[44] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2012.

[45] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol. 49, pp. 2415–2425, Oct 2003.

[46] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Transactions on Information Theory*, vol. 52, pp. 1522–1544, April 2006.

[47] J. Jiang, J. S. Thompson, P. M. Grant, and N. Goertz, "Practical compress-and-forward cooperation for the classical relay network," in *2009 17th European Signal Processing Conference*, pp. 2421–2425, Aug 2009.

[48] M. Kuhn, J. Wagner, and A. Wittneben, "Cooperative processing for the WLAN uplink," in *2008 IEEE Wireless Communications and Networking Conference*, pp. 1294–1299, March 2008.

[49] C. X. Wang, X. Hong, X. Ge, X. Cheng, G. Zhang, and J. Thompson, "Cooperative MIMO channel models: A survey," *IEEE Communications Magazine*, vol. 48, pp. 80–87, February 2010.

[50] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, pp. 80–89, Sept 2004.

[51] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, pp. 8–27, Aug 2002.

[52] X. Hong, Y. Jie, C. X. Wang, J. Shi, and X. Ge, "Energy-spectral efficiency trade-off in virtual MIMO cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 2128–2140, October 2013.

[53] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 2020–2040, June 2005.

[54] J. Jiang, J. S. Thompson, H. Sun, and P. M. Grant, "Performance assessment of virtual multiple-input multiple-output systems with compress-and-forward cooperation," *IET Communications*, vol. 6, pp. 1456–1465, July 2012.

[55] J. Jiang, M. Dianati, M. A. Imran, R. Tafazolli, and S. Zhang, "Energy-efficiency analysis and optimization for virtual-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 2272–2283, Jun 2014.

[56] C. V. N. Index, "Global mobile data traffic forecast update, 2016-2021," 2016.

[57] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, pp. 366–385, March 2014.

[58] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proceedings of the IEEE*, vol. 99, pp. 1390–1436, Aug 2011.

[59] Z. Qingling and J. Li, "Rain attenuation in millimeter wave ranges," in *2006 7th International Symposium on Antennas, Propagation EM Theory*, pp. 1–4, Oct 2006.

[60] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[61] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan 2013.

[62] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, pp. 403–430, Jan 2017.

[63] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys Tutorials*, vol. 18, pp. 949–973, Secondquarter 2016.

[64] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, pp. 106–113, February 2014.

[65] S. Geng, J. Kivinen, X. Zhao, and P. Vainikainen, "Millimeter-wave propagation channel characterization for short-range wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 3–13, Jan 2009.

[66] R. C. Daniels, J. N. Murdock, T. S. Rappaport, and R. W. Heath, "60 GHz wireless: Up close and personal," *IEEE Microwave Magazine*, vol. 11, pp. 44–50, Dec 2010.

[67] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall communications engineering and emerging technologies series, Dorling Kindersley, 2009.

[68] S.-K. Yong, P. Xia, and A. Valdes-Garcia, *60GHz Technology for Gbps WLAN and WPAN: From Theory to Practice*. Wiley Publishing, 2011.

[69] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5G: opportunities and challenges," *Wireless Networks*, vol. 21, pp. 2657–2676, Nov 2015.

[70] K. Zheng, L. Zhao, J. Mei, M. Dohler, W. Xiang, and Y. Peng, "10 Gb/s HetsNets with millimeter-wave communications: access and networking - challenges and protocols," *IEEE Communications Magazine*, vol. 53, pp. 222–231, January 2015.

[71] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, pp. 3122–3134, November 2011.

[72] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," *CoRR*, vol. abs/1604.03183, 2016.

[73] F. Baccelli and B. Bllaszczyszyn, "Stochastic geometry and wireless networks: Volume I theory," *Found. Trends Netw.*, vol. 3, pp. 249–449, Mar. 2009.

[74] F. Baccelli and B. Bllaszczyszyn, "Stochastic geometry and wireless networks: Volume II applications," *Found. Trends Netw.*, vol. 4, pp. 1–312, Jan. 2009.

[75] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*. Springer Series in Statistics, Springer New York, 2013.

[76] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 1029–1046, September 2009.

[77] A. He, D. Liu, Y. Chen, and T. Zhang, "Stochastic geometry analysis of energy efficiency in hetnets with combined CoMP and BS sleeping," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 1798–1802, Sept 2014.

[78] D. P. Kroese and Z. I. Botev, "Spatial process simulation," in *Stochastic Geometry, Spatial Statistics and Random Fields*, pp. 369–404, Springer, 2015.

[79] S. B. Lowen and M. C. Teich, "Power-law shot noise," *IEEE Transactions on Information Theory*, vol. 36, pp. 1302–1318, Nov 1990.

[80] T. X. Brown, "Cellular performance bounds via shotgun cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 2443–2455, Nov 2000.

[81] F. Baccelli and S. Zuyev, "Frontiers in queueing," ch. Stochastic Geometry Models of Mobile Communication Networks, pp. 227–243, Boca Raton, FL, USA: CRC Press, Inc., 1997.

[82] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, March 1982.

[83] Z. Ding, K. K. Leung, D. L. Goeckel, and D. Towsley, "Cooperative transmission protocols for wireless broadcast channels," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3701–3713, December 2010.

[84] J. Jiang, M. Dianati, M. A. Imran, and Y. Chen, "Energy efficiency and optimal power allocation in virtual-MIMO systems," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–6, Sept 2012.

[85] J. Jiang, J. S. Thompson, H. Sun, and P. M. Grant, "Practical analysis of codebook design and frequency offset estimation for virtual-multiple-input-multiple-output systems," *IET Communications*, vol. 7, pp. 585–594, April 2013.

[86] Y. Qi, R. Hoshyar, and R. Tafazolli, "A novel quantization scheme in compress-and-forward relay system," in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, pp. 1–5, April 2009.

[87] W. Jiang, F. Ding, and Q.-L. Xiang, "An affinity propagation based method for vector quantization codebook design," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, Dec 2008.

[88] P. Y. Yang, J. T. Tsai, and J. H. Chou, "PCA-based fast search method using PCA-LBG-Based VQ codebook for codebook search," *IEEE Access*, vol. 4, pp. 1332–1344, 2016.

[89] B. Hui and K. H. Chang, "Low complexity vector quantization algorithm for codebook optimization," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 868–872, April 2012.

[90] S. Zhang, Q. F. Zhou, C. Kai, and W. Zhang, "Channel quantization based physical-layer network coding," in *2013 IEEE International Conference on Communications (ICC)*, pp. 5137–5142, June 2013.

[91] P. Wu, L. Li, and P. Zhang, "Universal unitary space vector quantization codebook design for precoding MIMO system under spatial correlated channel," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, pp. 435–439, May 2008.

[92] C. K. Au-yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 458–462, Feb 2007.

[93] M. F. Jian, W. J. Huang, and C. K. Wen, "Distributed beamforming with compressed feedback in time-varying cooperative networks," in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–4, Dec 2012.

[94] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Transactions on Signal Processing*, vol. 54, pp. 1853–1863, May 2006.

[95] S. M. Satti, N. Deligiannis, A. Munteanu, P. Schelkens, and J. Cornelis, "Symmetric scalable multiple description scalar quantization," *IEEE Transactions on Signal Processing*, vol. 60, pp. 3628–3643, July 2012.

[96] T. K. Kim, W. Choi, and G. H. Im, "Efficient codebook design for co-operative MIMO systems with decode-and-forward relay," *IEEE Communications Letters*, vol. 20, pp. 598–601, March 2016.

[97] X. Ma, Z. Pan, Y. Li, and J. Fang, "High-quality initial codebook design method of vector quantisation using grouping strategy," *IET Image Processing*, vol. 9, no. 11, pp. 986–992, 2015.

[98] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating decoders," *IEEE Transactions on Information Theory*, vol. 52, pp. 5438–5454, Dec 2006.

[99] R. Dabora and S. D. Servetto, "On the role of estimate-and-forward with time sharing in cooperative communication," *IEEE Transactions on Information Theory*, vol. 54, pp. 4409–4431, Oct 2008.

[100] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, pp. 2325–2383, Oct 1998.

[101] J. Jiang, S. T. John, and M. G. Peter, "Design and analysis of compress-and-forward cooperation in a virtual-MIMO detection system," in *2010 IEEE Globecom Workshops*, pp. 126–130, Dec 2010.

[102] C. T. K. Ng, N. Jindal, A. J. Goldsmith, and U. Mitra, "Capacity gain from two-transmitter and two-receiver cooperation," *IEEE Transactions on Information Theory*, vol. 53, pp. 3822–3827, Oct 2007.

[103] P. Fertl, J. Jalden, and G. Matz, "Performance assessment of MIMO-BICM demodulators based on mutual information," *IEEE Transactions on Signal Processing*, vol. 60, pp. 1366–1382, March 2012.

[104] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient MMSE-based demodulator for MIMO bit-interleaved coded modulation," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 4, Nov 2004.

[105] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Kluwer international series in engineering and computer science: Communications and information theory, Springer US, 1992.

[106] R. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, pp. 4–29, April 1984.

[107] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, pp. 84–95, Jan 1980.

[108] Y. You, *Audio Coding: Theory and Applications*. Springer US, 2010.

[109] A. Agarwal and A. K. Jagannatham, "Optimal wake-up scheduling for PSM delay minimization in mobile wireless networks," *IEEE Wireless Communications Letters*, vol. 2, pp. 419–422, August 2013.

[110] P. Scheunders, "A genetic lloyd-max image quantization algorithm," *Pattern Recognition Letters*, vol. 17, no. 5, pp. 547 – 556, 1996.

[111] J. Shanbehzadeh and P. O. Ogunbona, "On the computational complexity of the LBG and PNN algorithms," *IEEE Transactions on Image Processing*, vol. 6, pp. 614–616, Apr 1997.

[112] S. Tombaz, A. Vastberg, and J. Zander, "Energy- and cost-efficient ultra-high-capacity wireless access," *IEEE Wireless Communications*, vol. 18, pp. 18–24, October 2011.

[113] Y. Pei, T.-H. Pham, and Y. C. Liang, "How many RF chains are optimal for large-scale MIMO systems when circuit power is considered?," in *2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 3868–3873, Dec 2012.

[114] J. Gomez-Vilardebo, A. I. Perez-neira, and M. Najar, "Energy efficient communications over the AWGN relay channel," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 32–37, January 2010.

[115] Y. Yao, X. Cai, and G. B. Giannakis, "On energy efficiency and optimum resource allocation of relay transmissions in the low-power regime," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2917–2927, Nov 2005.

[116] S. K. Jayaweera, "Virtual MIMO-based cooperative communication for energy-constrained wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 984–989, May 2006.

[117] L. Gerdes, L. Weiland, and W. Utschick, "A zero-forcing partial decode-and-forward scheme for the gaussian MIMO relay channel," in *2013 IEEE International Conference on Communications (ICC)*, pp. 3349–3354, June 2013.

[118] M. T. Kakitani, G. Brante, R. D. Souza, and M. A. Imran, "Energy efficiency of transmit diversity systems under a realistic power consumption model," *IEEE Communications Letters*, vol. 17, pp. 119–122, January 2013.

[119] G. Miao, "Energy-efficient uplink multi-user MIMO," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2302–2313, May 2013.

[120] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *2010 Future Network Mobile Summit*, pp. 1–8, June 2010.

[121] A. R. Jensen, M. Lauridsen, P. Mogensen, T. B. Srensen, and P. Jensen, "LTE UE power consumption model: For system level energy and performance optimization," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sept 2012.

[122] D. Ha, K. Lee, and J. Kang, "Energy efficiency analysis with circuit power consumption in massive MIMO systems," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 938–942, Sept 2013.

[123] E. Bjrnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 242–247, April 2014.

[124] G. L. Stuber, J. R. Barry, S. W. McLaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband mimo-ofdm wireless communications," *Proceedings of the IEEE*, vol. 92, pp. 271–294, Feb 2004.

[125] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?," *IEEE Transactions on Communications*, vol. 63, pp. 3917–3928, Nov 2015.

[126] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 293–298, Oct 2013.

[127] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[128] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3590–3600, November 2010.

[129] L. Zhao, H. Zhao, F. Hu, K. Zheng, and J. Zhang, "Energy efficient power allocation algorithm for downlink massive MIMO with MRT precoding," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sept 2013.

[130] L. Liu, Y. Yi, J. F. Chamberland, and J. Zhang, "Energy-efficient power allocation for delay-sensitive multimedia traffic over wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 2038–2047, Jun 2014.

[131] R. S. Prabhu and B. Daneshrad, "Performance analysis of energy-efficient power allocation for MIMO-MRC systems," *IEEE Transactions on Communications*, vol. 60, pp. 2048–2053, August 2012.

[132] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Transactions on Vehicular Technology*, vol. 64, pp. 5275–5287, Nov 2015.

[133] L. Sboui, Z. Rezki, and M. S. Alouini, "On energy efficient power allocation for power-constrained systems," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 1954–1958, Sept 2014.

[134] R. Vaca-Ramirez, *Low complexity radio resource management for energy efficient wireless networks*. PhD thesis, The University of Edinburgh, 2014.

[135] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4298–4310, Aug 2010.

[136] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Utilizing the spatial information provided by channel norm feedback in SDMA systems," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3278–3293, July 2008.

[137] F. Boccardi and H. Huang, "Optimum power allocation for the MIMO-BC zero-forcing precoder with per-antenna power constraints," in *2006 40th Annual Conference on Information Sciences and Systems*, pp. 504–504, March 2006.

[138] E. Björnson, E. Jorswieck, *et al.*, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2–3, pp. 113–381, 2013.

[139] J. Hoydis, S. ten Brink, and M. Debbah, "Comparison of linear precoding schemes for downlink massive MIMO," in *2012 IEEE International Conference on Communications (ICC)*, pp. 2135–2139, June 2012.

[140] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Communications*, vol. 22, pp. 13–21, October 2015.

[141] M. Mirahsan, R. Schoenen, H. Yanikomeroglu, G. Senarath, and N. Dung-Dao, "User-in-the-loop for hethetnets with backhaul capacity constraints," *IEEE Wireless Communications*, vol. 22, pp. 50–57, October 2015.

[142] T. Zhang, J. Zhao, L. An, and D. Liu, "Energy efficiency of base station deployment in ultra dense HetNets: A stochastic geometry analysis," *IEEE Wireless Communications Letters*, vol. 5, pp. 184–187, April 2016.

[143] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: a 5G perspective," *IEEE Communications Magazine*, vol. 52, pp. 66–73, February 2014.

[144] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, pp. 4391–4403, October 2013.

[145] C. Dehos, J. L. Gonzalez, A. D. Domenico, D. Ktenas, and L. Dussopt, "Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?," *IEEE Communications Magazine*, vol. 52, pp. 88–95, September 2014.

[146] H. Wang, J. Ding, Y. Li, P. Hui, J. Yuan, and D. Jin, "Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks," in *Proceedings of the 7th International Workshop on Hot Topics in Planet-scale mObile Computing and Online Social neTworking*, HOTPOST '15, (New York, NY, USA), pp. 19–24, ACM, 2015.

[147] C. Saha and H. S. Dhillon, "Downlink coverage probability of k-tier HetNets with general non-uniform user distributions," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2016.

[148] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "HetHetNets: Heterogeneous traffic distribution in heterogeneous wireless cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 2252–2265, Oct 2015.

[149] M. Mirahsan, R. Schoenen, S. S. Szyszkowicz, and H. Yanikomeroglu, "Measuring the spatial heterogeneity of outdoor users in wireless cellular networks based on open urban maps," in *2015 IEEE International Conference on Communications (ICC)*, pp. 2834–2838, June 2015.

[150] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Modeling non-uniform UE distributions in downlink cellular networks," *IEEE Wireless Communications Letters*, vol. 2, pp. 339–342, June 2013.

[151] M. Mirahsan, H. Yanikomeroglu, G. Senarath, and N. D. Dao, "Analytic modeling of SIR in cellular networks with heterogeneous traffic," *IEEE Communications Letters*, vol. 20, pp. 1627–1630, Aug 2016.

[152] C. Saha, M. Afshang, and H. S. Dhillon, "Enriched k-tier HetNet model to enable the analysis of user-centric small cell deployments," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 1593–1608, March 2017.

[153] C. C. Coskun and E. Ayanoglu, "A greedy algorithm for energy-efficient base station deployment in heterogeneous networks," in *2015 IEEE International Conference on Communications (ICC)*, pp. 7–12, June 2015.

[154] K. Son, E. Oh, and B. Krishnamachari, "Energy-efficient design of heterogeneous cellular networks from deployment to operation," *Computer Networks*, vol. 78, pp. 95 – 106, 2015. Special Issue: Green Communications.

[155] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 840–850, May 2013.

[156] K. Son, E. Oh, and B. Krishnamachari, "Energy-aware hierarchical cell configuration: From deployment to operation," in *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 289–294, April 2011.

[157] A. Zappone, E. Bjrnson, L. Sanguinetti, and E. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks," *IEEE Transactions on Signal Processing*, vol. 65, pp. 2844–2859, June 2017.

[158] M. Feldman, J. Naor, and R. Schwartz, "A unified continuous greedy algorithm for submodular maximization," in *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pp. 570–579, Oct 2011.

[159] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd ed., 2001.

[160] A. Krause and D. Golovin, *Submodular Function Maximization*, pp. 71–104. Cambridge University Press, 2014.

[161] A. Vince, "A framework for the greedy algorithm," *Discrete Applied Mathematics*, vol. 121, no. 1, pp. 247 – 260, 2002.

[162] A. Clark, B. Alomair, L. Bushnell, R. Poovendran, and J. Baillieul, *Submodularity in Dynamics and Control of Networked Systems*. Communications and Control Engineering, Springer International Publishing, 2015.

[163] F. Bach *et al.*, "Learning with submodular functions: A convex optimization perspective," *Foundations and Trends® in Machine Learning*, vol. 6, no. 2-3, pp. 145–373, 2013.

[164] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, pp. 1850–1859, April 2013.

[165] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, pp. 831–846, Oct 2014.

[166] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution?," *IEEE Communications Magazine*, vol. 52, pp. 56–62, September 2014.

[167] B. Badic, T. O'Farrrell, P. Loskot, and J. He, "Energy efficient radio access architectures for green radio: Large versus small cell size deployment," in *2009 IEEE 70th Vehicular Technology Conference Fall*, pp. 1–5, Sept 2009.

[168] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, pp. 56–61, June 2011.

[169] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, "Energy-efficient base-station sleep-mode techniques in green cellular networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 803–826, Secondquarter 2015.

[170] G. Lim and L. J. Cimini, "Energy efficiency of cooperative beamforming in wireless ad-hoc networks," in *2012 IEEE International Conference on Communications (ICC)*, pp. 4039–4043, June 2012.

[171] J. Qiao, B. Cao, X. Zhang, X. Shen, and J. W. Mark, "Efficient concurrent transmission scheduling for cooperative millimeter wave systems," in *2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 4187–4192, Dec 2012.

[172] R. Subbu and A. Sanderson, *Network-based Distributed Planning Using Coevolutionary Algorithms*. Series in intelligent control and intelligent automation, World Scientific Publishing Company, 2004.

[173] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, 2015.

[174] H. S. Dhillon and J. G. Andrews, "Downlink rate distribution in heterogeneous cellular networks under generalized cell selection," *IEEE Wireless Communications Letters*, vol. 3, pp. 42–45, February 2014.

[175] M. Afshang, H. S. Dhillon, and P. H. J. Chong, "Modeling and performance analysis of clustered device-to-device networks," *CoRR*, vol. abs/1508.02668, 2015.

[176] M. Afshang, H. S. Dhillon, and P. H. J. Chong, "Fundamentals of cluster-centric content placement in cache-enabled device-to-device networks," *IEEE Transactions on Communications*, vol. 64, pp. 2511–2526, June 2016.

[177] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 550–560, April 2012.

[178] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 3484–3495, October 2012.

[179] Q. Shi, C. He, H. Chen, and L. Jiang, "Distributed wireless sensor network localization via sequential greedy optimization algorithm," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3328–3340, June 2010.

[180] J. H. Lee and W. Choi, "Unified codebook design for vector channel quantization in MIMO broadcast channels," *IEEE Transactions on Signal Processing*, vol. 63, pp. 2509–2519, May 2015.

[181] D. Kern and V. Kuehn, "On information bottleneck graphs to design compress and forward quantizers with side information for multi-carrier transmission," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.